INTRODUCCIÓN AL CÁLCULO Y AL ANÁLISIS MATEMÁTICO

INTRODUCCIÓN AL CÁLCULO Y AL ANÁLISIS MATEMÁTICO

Vol. I

RICHARD COURANT y FRITZ JOHN

DEL COURANT INSTITUTE OF MATHEMATICAL SCIENCES, UNIVERSIDAD DE NUEVA YORK



LIMUSA

NORIEGA EDITORES

MÉXICO • España • Venezuela • Colombia

VERSIÓN AUTORIZADA EN ESPAÑOL DE LA OBRA PUBLICADA EN INGLÉS CON EL TÍTULO:
INTRODUCTION TO CALCULUS AND ANALYSIS. VOLUME 1
© JOHN WILEY & SONS, INC.

COLABORADORES EN LA TRADUCCIÓN:
SAÚL HAHN GOLDBERG
INGENIERO EN COMUNICACIONES Y ELECTRÓNICA. LICENCIADO EN CIENCIAS FISICOMATEMÁTICAS Y PROFESOR DE MATEMÁTICAS EN LA ESCUELA SUPERIOR DE FÍSICA Y MATEMÁTICAS DEL INSTITUTO POLITÉCNICO NACIONAL, MÉXICO.

ROLANDO V. JIMÉNEZ DOMÍNGUEZ
INGENIERO EN COMUNICACIONES Y ELECTRÓNICA. LICENCIADO EN FISICOMATEMÁTICAS Y PROFESOR DE FÍSICA EN LA ESCUELA SUPERIOR DE FÍSICA Y MATEMÁTICAS DEL INSTITUTO POLITÉCNICO NACIONAL, MÉXICO.

JOSÉ S. FLORIO
PROFESOR Y DOCTOR EN CIENCIAS MATEMÁTICAS. JEFE DE LA DIVISIÓN DE CIENCIAS Y DIRECTOR DEL INSTITUTO TECNOLÓGICO, UNIVERSIDAD DE LAS AMÉRICAS.

BALDERAS 95, MÉXICO, D.F.
C.P. 06040
☎ (5)521-21-05
01(800) 7-06-91-00
(5)512-29-03
limusa@noriega.com.mx
www.noriega.com.mx

CANIEM NÚM. 121

DECIMOSEXTA REIMPRESIÓN

IMPRESO EN MÉXICO
ISBN 968-18-0639-5



Prólogo

En la segunda mitad del siglo XVIII el nuevo análisis matemático surgió como la fuerza dominante en la Matemática. Este análisis se caracteriza por la extraordinariamente eficaz operación que se funda en procesos infinitos, o sea, límites. Dos de estos procesos, la derivación y la integración, se convirtieron en la esencia de la disciplina, llamada Cálculo Diferencial e Integral o, a menudo, simplemente "Cálculo", que es la base de *todo* el análisis matemático.

Se apreció, de inmediato, la importancia de los nuevos descubrimientos y métodos, los cuales causaron una profunda *conmoción intelectual*. Sin embargo, el dominio de este poderoso arte matemático constituyó, al principio, un objetivo formidable y difícil de lograr, ya que los textos sobre la materia eran escasos, poco sistemáticos y, con frecuencia carentes de claridad. Por tanto, fue venturoso para la Matemática y para la Ciencia, en general, que las figuras principales del nuevo movimiento reconocieran pronto la vital necesidad de escribir tratados con el fin de que la materia fuera accesible a un público mucho más numeroso que el pequeño círculo intelectual de esa época. Uno de los matemáticos más destacados de los tiempos modernos, Leonhard Euler, estableció los fundamentos de una larga tradición en sus libros introductorios; y aunque se ha progresado mucho en la aclaración y simplificación del material del Cálculo, estos libros del siglo XVIII, en la actualidad aún son fuente de inspiración.

Después de Euler, los autores, uno tras otro, se solidarizaron con la separación entre cálculo diferencial y cálculo integral y, al hacerlo, oscurecieron un punto clave: la reciprocidad entre derivación e integración. Fue en 1927, cuando apareció, publicada por "Springer-Verlag", la primera edición alemana de la obra de R. Courant "Vorlesungen über Differential and Integralrechnung", que se eliminó tal separación y el Cálculo se presentó como una disciplina unificada.

La presente obra se originó en aquel primer libro alemán y sus ediciones subsiguientes. A partir de 1939, con la colaboración de James y Virginia

McShane, la casa "Blackie & Sons" de Glasgow, preparó y publicó una versión inglesa muy ampliada y modificada, de la cual la Interscience-Wiley distribuyó numerosas reimpresiones en los Estados Unidos de Norteamérica.

Con el transcurso de los años, se hizo evidente que los programas de la enseñanza superior en los Estados Unidos, requerían una nueva revisión de la obra. Por otra parte, parecía poco prudente alterar las versiones originales que todavía son útiles.

Así, en vez de modificar la obra existente, se optó por complementarla con un libro esencialmente nuevo, relacionado en muchos sentidos con los originales europeos pero, en términos más específicos, orientado para satisfacer las necesidades de la actual y futuras generaciones de estudiantes en los Estados Unidos. Tal plan resultó factible cuando Fritz John, que ya había ayudado mucho en la preparación de la primera edición inglesa, convino en escribir de nuevo el libro con R. Courant.

Aunque este libro difiere notablemente del original, tanto en forma como en contenido, tiene el mismo propósito: guiar al estudiante directamente a la esencia del tema y capacitarlo para aplicar sus conocimientos. La obra evita el estilo dogmático que oculta la motivación de los conceptos y las raíces que el Cálculo tiene en la realidad intuitiva. Un importante objetivo que se persigue en este libro es mostrar la relación recíproca entre el análisis matemático y sus diversas aplicaciones y destacar el papel de la intuición. Esperamos que cierto énfasis en la precisión no interfiera con este objetivo.

La Matemática presentada como un sistema de verdades, acabado y ordenado, sin referencia al origen y propósito de sus conceptos y teorías, tiene su encanto y satisface una necesidad filosófica. Pero esa actitud introvertida en el campo de la ciencia, no es adecuada para los estudiantes que buscan independencia intelectual, más bien que indoctrinación. Y menospreciar las aplicaciones e intuición conduce al aislamiento y atrofia en la Matemática. Entonces, resulta sumamente importante que estudiantes y maestros estén a salvo del purismo presumido.

El libro está dirigido a estudiantes de varios niveles, a matemáticos, científicos en general e ingenieros. No pretende facilitar la materia suavizando las dificultades, sino que, más bien, trata de ayudar al lector sinceramente interesado, a base de aclarar las interrelaciones y propósitos de la obra en conjunto.

En lugar de obstaculizar el acceso al caudal de hechos y propiedades por medio de largas exposiciones de carácter fundamental, éstas se han pospuesto, a veces, relegándolas a apéndices de los diversos capítulos.

Se dan numerosos ejemplos y problemas al final de los capítulos. Algunos son fascinantes, otros llegan a ser verdaderamente difíciles; y la mayoría de ellos complementan el material teórico del texto. Se preparará un manual complementario de esta obra con problemas y ejercicios de carác-

ter más rutinario, y, además, se agregarán respuestas y sugerencias para resolverlos.

Muchos colegas y amigos han ayudado. Albert A. Blank no sólo aportó críticas perspicaces y constructivas, sino que también desempeñó un papel importante en la labor de ordenar, aumentar y seleccionar los problemas y ejercicios y, además, asumió la responsabilidad principal de preparar el manual mencionado.* Alan Solomon ayudó, muy desinteresada y eficazmente, en todas las etapas de la preparación del libro. También se les agradece a Charlotte John, Anneli Lax, R. Richtmyer, y a otros amigos, entre los que figuran James y Virginia McShane.

Este volumen trata principalmente de funciones de una sola variable, mientras que en el Volumen II se estudiarán las teorías del Cálculo más diversificadas para funciones de varias variables. Al lector se le debe hacer una observación final, ya que podría resultar un fracaso tratar de dominar la materia estudiando un libro de este tipo página por página, sin desviarse del camino trazado. Se puede alcanzar gradualmente una mejor comprensión del tema y una visión más amplia, *sólo tomando atajos* en la primera lectura y, luego, volviendo repetidas veces a los mismos temas, preguntas y dificultades.

Se ha tratado de ayudar al lector señalando con asteriscos ciertos tópicos que podrían causarle trastornos al estudiarlos por primera vez, así como algunos de los problemas más difíciles.

Esperamos que la obra, con este nuevo enfoque, será útil a la joven generación de científicos. Estamos conscientes de sus muchas imperfecciones, y sinceramente solicitamos comentarios críticos que podrían ser beneficiosos para mejoras posteriores.

RICHARD COURANT
FRITZ JOHN

* Manual de Problemas de Cálculo y Análisis Matemático, por Albert A. Blank, publicado en español por la Editorial Limusa

Contenido

INTRODUCCIÓN AL CÁLCULO Y AL ANÁLISIS MATEMÁTICO

1

Introducción

Desde la antigüedad las nociones intuitivas de cambio continuo, crecimiento y movimiento, han atraído la atención de las mentes científicas. Empero, el camino hacia el entendimiento de variación continua no fue logrado sino hasta el siglo XVII, cuando la ciencia moderna surgió y se desarrolló rápidamente en cercana conjunción con el cálculo integral y diferencial, llamado brevemente Cálculo, y el análisis matemático.

Las nociones básicas del Cálculo son las de derivada e integral: la derivada es una medida de la "rapidez" de una variación; la integral, una medida del efecto total de un proceso de cambio continuo. Una comprensión precisa de estos conceptos y de su abrumadora fecundidad descansa sobre los conceptos de límite y función; los cuales a su vez dependen de un entendimiento del continuo de números. Sólo gradualmente, penetrando más y más en la esencia del Cálculo, puede uno apreciar su potencia y belleza. En este capítulo introductorio se explicarán los conceptos básicos de número, función y límite, primero de manera simple e intuitiva, y después mediante argumentos cuidadosos.

1.1 El continuo de números

Los enteros positivos o *números naturales* 1, 2, 3, ... son símbolos abstractos para indicar "cuántos" objetos hay en una *colección* o *conjunto* de elementos discretos.

Estos símbolos están despojados de toda referencia a las cualidades concretas de los objetos contados, ya sea que se trate de personas, átomos, casas o cualesquiera objetos.

Los números naturales son el instrumento adecuado para contar elementos de una colección o "conjunto". Sin embargo, no bastan para otro objetivo igualmente importante: *medir* cantidades tales como la longitud de una curva y el volumen o peso de un cuerpo. La pregunta "¿cuánto?"

no puede ser contestada inmediatamente en términos de los números naturales. La profunda necesidad de expresar medidas de cantidades en términos de lo que nosotros quisiéramos llamar números nos obliga a extender el concepto de número de manera que podamos describir una graduación continua de las medidas. Esta extensión es llamada el *continuo de números* o el sistema de "números reales" (un nombre no descriptivo pero aceptado generalmente). La extensión del concepto de número al de continuo es de tal manera convincente que fue utilizada por todos los grandes matemáticos y científicos de épocas anteriores sin hacer preguntas escudriñadoras. No fue sino hasta el siglo XIX que los matemáticos se vieron impulsados a buscar un fundamento lógico más firme para el sistema de los números reales. La subsiguiente formulación precisa de estos conceptos condujo, a su vez, a un progreso mayor en matemáticas. Principiaremos con un tratamiento intuitivo sin complicaciones, y posteriormente se dará un análisis más profundo del sistema de los números reales.[1]

a. El sistema de números naturales y su extensión. Numeración y medición

Los números naturales y los racionales. La sucesión de números "naturales" 1, 2, 3, . . . se considera como dada. No se requiere discutir cómo estos entes abstractos, los números, pueden ser catalogados desde un punto de vista filosófico. Para el matemático, y para cualquiera que trabaje con números, es importante conocer solamente las reglas o leyes mediante las cuales éstos pueden combinarse para obtener otros números naturales. Estas leyes forman la base de las reglas conocidas para sumar y multiplicar números en el sistema decimal; incluyen las *leyes conmutativas* $a + b = b + a$ y $ab = ba$, las *leyes asociativas* $a + (b + c) = (a + b) + c$ y $a(bc) = (ab)c$, la *ley distributiva* $a(b + c) = ab + ac$, la *ley de la cancelación,* esto es que $a + c = b + c$ implica $a = b$, etc.

Las operaciones inversas, substracción y división, no son siempre posibles dentro del conjunto de los números naturales; no podemos restar 2 de 1 o dividir 1 entre 2 y permanecer dentro de ese conjunto. Para hacer posibles estas operaciones sin restricción, nos vemos obligados a extender el concepto de número inventando el número 0, los enteros "negativos" y las fracciones. La totalidad de todos estos números se llama la clase o conjunto de *números racionales;* todos ellos se obtienen de la unidad utilizando las "operaciones racionales" de cálculo, a saber: la adición, substracción, multiplicación y división.[2]

---

[1] Una exposición más completa se da en *What Is Mathematics?* por Courant y Robbins, Oxford University Press, 1962.

[2] Aquí, la palabra "racional" no significa razonable o lógico sino que se deduce de la palabra "razón", significando la proporción relativa de dos magnitudes.

Un número racional puede expresarse siempre en la forma p/q, donde p y q son enteros y $q \neq 0$. Podemos hacer que esta representación sea única exigiendo que q sea positivo y que p y q no tengan un factor común mayor que 1.

Dentro del dominio de los números racionales todas las *operaciones racionales,* adición, multiplicación, substracción y división (excepto la división por cero), pueden efectuarse y producir nuevamente números racionales. Como se sabe de la aritmética elemental, las operaciones con números racionales obedecen las mismas leyes que las operaciones con números naturales; por ello los números racionales extienden el sistema de enteros positivos en una forma completamente natural.

Representación gráfica de los números racionales. Los números racionales son usualmente representados en forma gráfica mediante puntos en una línea recta L, el *eje numérico* (o *recta numérica).* Tomando un

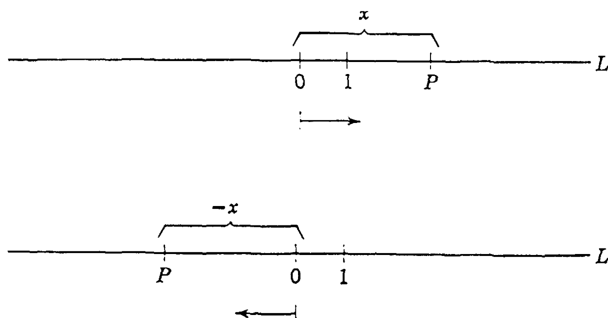


Figura 1.1 El eje numérico.

punto arbitrario de L como el origen o punto 0 y otro punto arbitrario como el punto 1, utilizamos la distancia entre estos dos puntos como escala o unidad de medida y definimos la dirección de 0 a 1 como "positiva". La línea recta con una dirección así impuesta se llama recta dirigida. Es costumbre dibujar a L de manera que el punto 1 se encuentre a la derecha del punto 0 (Fig. 1.1). La localización de cualquier punto P en L está completamente determinada por dos partes de información: la distancia a P desde el origen 0 y la dirección de 0 a P (a la derecha o izquierda del 0). El punto P en L que representa un número racional positivo se encuentra a una distancia de x unidades a la derecha del 0. Un número racional negativo x es representado por el punto $-x$ unidades a la izquierda del 0. En cualquier caso la distancia del 0 al punto que representa a x se llama el valor absoluto de x, denotado por $|x|$, y se tiene

$$|x| = \begin{cases} x, & \text{si } x \text{ es positivo o cero,} \\ -x, & \text{si } x \text{ es negativo} \end{cases}$$

Se observa que x nunca es negativo y es igual a cero sólo cuando $x = 0$.

Recordemos de la geometría elemental que con regla y compás es posible subdividir la unidad de longitud en un número cualquiera de partes iguales. Se sigue que cualquier longitud racional puede ser construida, y, por lo tanto, que el punto que representa un número racional x puede ser encontrado por métodos puramente geométricos.

De esta forma se obtiene una representación geométrica de números racionales mediante puntos en L, los llamados *puntos racionales.* Conforme a nuestra notación para los puntos 0 y 1, nos tomamos la libertad de denotar tanto los números racionales como los puntos correspondientes en L por el mismo símbolo x.

La relación $x < y$ para dos números racionales significa geométricamente que el punto x se encuentra a la izquierda del punto y. En ese caso la distancia entre los puntos es $y - x$ unidades. Si $x > y$, la distancia es $x - y$ unidades. En cualquier caso la distancia entre dos puntos racionales x, y de L es $|y - x|$ unidades y es también un número racional.

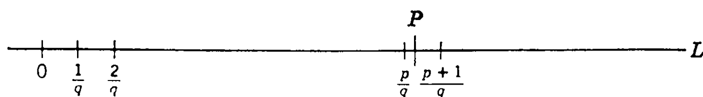


Figura 1.2

Un segmento en L con puntos extremos a, b con $a < b$ se denominará *intervalo.* El segmento particular con puntos extremos 0, 1 se llama *intervalo unitario.* Si los puntos extremos están incluidos en el intervalo, se dirá que el intervalo es *cerrado;* si los puntos extremos se excluyen, el intervalo se llama *abierto.* El intervalo abierto, denotado por (a, b), consiste en aquellos puntos x para los cuales $a < x < b$, esto es, en aquellos puntos que se encuentran "entre" a y b. El intervalo cerrado, denotado por $[a, b]$, consiste en los puntos x para los cuales $a < x < b$.[1] En cualquier caso la *longitud del intervalo* es $b - a$.

Los puntos que corresponden a los enteros 0, ± 1, ± 2, ... subdividen el eje numérico en intervalos de longitud unitaria. Todo punto en L es un punto extremo o bien un punto interior de uno de los intervalos de la subdivisión. Si se sigue subdividiendo todo intervalo en q partes iguales, se obtiene una subdivisión de L en intervalos de longitud $1/q$ por puntos racionales de la forma p/q. Todo punto P de L es entonces un punto racional de la forma p/q, o bien se encuentra entre dos puntos racionales sucesivos p/q y $(p + 1)/q$ (ver Fig. 1.2). Puesto que puntos sucesivos de la subdivisión distan en $1/q$ unidades, se sigue que podemos encontrar un punto

[1] La relación $a \leq x$ (léase "a es menor que o igual a x") es interpretada como "$a < x$, o bien $a = x$". Los dobles signos $\geq$ y $\pm$ se interpretan de manera análoga.

racional p/q cuya distancia desde P no excede $1/q$ unidades. El número $1/q$ puede hacerse tan pequeño como se desee escogiendo q como un entero positivo suficientemente grande. Por ejemplo, escogiendo $q = 10^n$ (donde n es cualquier número natural), podemos encontrar una "fracción decimal" $x = p/10^n$ cuya distancia desde P es menor que $1/10^n$. Aunque no se afirma que todo punto de L es un punto racional, se ve por lo menos que pueden encontrarse puntos racionales arbitrariamente próximos a cualquier punto P de L.

Densidad

La proximidad arbitraria de puntos racionales a un punto P dado de L se expresa diciendo: *Los puntos racionales son densos en el eje numérico.* Es claro que aun conjuntos menores de números racionales son densos, por ejemplo, los puntos $x = p/10^n$, para todos los números naturales n y enteros p.

La densidad implica que entre dos puntos racionales cualesquiera a y b existe una infinidad de otros puntos racionales. En particular, el punto medio entre a y b, $c = \frac{1}{2}(a + b)$, correspondiente a la media aritmética de los números a y b, es nuevamente racional. Tomando los puntos medios de a y c, de b y c, y continuando de esta manera, se puede obtener cualquier número de puntos racionales entre a y b.

Un punto arbitrario P en L puede ser localizado con cualquier grado de precisión utilizando puntos racionales. Puede parecer entonces a primera vista que la labor de localizar P por un número ha sido lograda introduciendo los números racionales. Después de todo, en la realidad física las cantidades nunca se dan o conocen con absoluta precisión sino que siempre con sólo un grado de incertidumbre y por ello pueden muy bien ser consideradas como medidas por números racionales.

Cantidades inconmensurables. Los números racionales, aunque densos, no bastan como base teórica de medición por medio de números. Dos cantidades cuya razón es un número racional se llaman *conmensurables,* debido a que pueden ser expresadas como múltiplos enteros de una unidad común. Desde época tan remota como la de los siglos V o VI A.C. los matemáticos y filósofos griegos hicieron el sorprendente y profundo descubrimiento de que existen cantidades que no son conmensurables con una unidad dada. En particular, existen segmentos de recta que no son múltiplos racionales de un segmento unitario dado.

Es fácil dar un ejemplo de una longitud inconmensurable con la longitud unidad: la diagonal l de un cuadrado con lados de longitud unidad. Pues, por el teorema de Pitágoras, el cuadrado de esta longitud l debe ser

igual a 2. Por ello, si l fuese un número racional y consecuentemente igual a p/q, donde p y q son enteros positivos, tendríamos $p^2 = 2q^2$. Podemos suponer que p y q no tienen factores comunes, pues tales factores comunes podrían ser cancelados en primer lugar. De acuerdo con la ecuación anterior, p^2 es un número par; de aquí que p mismo debe ser par, digamos $p = 2p'$. Sustituyendo p por $2p'$ se obtiene $4p'^2 = 2q^2$, o bien $q^2 = 2p'^2$; consecuentemente, q^2 sería par y por lo tanto q también par. Esto es lo mismo que decir, que p y q tienen ambos el factor 2. Sin embargo, esto contradice nuestra hipótesis de que p y q no tienen un factor común. Puesto que la suposición de que la diagonal puede ser representada por una fracción p/q da lugar a una contradicción, la suposición es falsa.

Este razonamiento, ejemplo característico de *prueba indirecta*, muestra que el símbolo $\sqrt{2}$ no puede corresponder a ningún número racional. Otro ejemplo es π, la razón de la circunferencia de un círculo a su diámetro. La demostración de que π no es racional es mucho más complicada y se obtuvo solamente en tiempos modernos (Lambert, 1761). Es fácil encontrar muchas cantidades inconmensurables (véase problema 1, p. 129); de hecho, las cantidades inconmensurables son, en cierto sentido, mucho más comunes de las conmensurables (véase p. 122).

Números irracionales

Debido a que el sistema de números racionales no es suficiente para la geometría, es necesario inventar nuevos números como medidas de cantidades inconmensurables: estos nuevos números se llaman "irracionales". Los antiguos griegos no pusieron de relieve el concepto de número abstracto; sin embargo, consideraron entes geométricos, tales como segmentos de recta, como los elementos básicos. De manera puramente geométrica ellos desarrollaron un sistema lógico para tratar y operar con cantidades inconmensurables, así como también con las conmensurables (racionales). Este logro tan importante, iniciado por los pitagóricos, fue desarrollado grandemente por Eudoxio y está expresado extensamente en el famoso *Elementos* de Euclides. En los tiempos modernos las matemáticas fueron creadas de nuevo y vastamente desarrolladas sobre el fundamento de conceptos numéricos en lugar de conceptos geométricos. Con la introducción de la geometría analítica, se desarrolló una reversión de orientación e importancia en la relación antigua entre números y cantidades geométricas, y así la teoría clásica de los inconmensurables fue poco menos que olvidada u omitida. Se supuso como un hecho el que a todo punto del eje numérico le corresponde un número racional o irracional y que esta totalidad de números "reales" obedece las mismas leyes aritméticas que los números racionales. Sólo posteriormente, en el siglo XIX, se sintió la necesidad de justificar

tal hipótesis y finalmente fue satisfecha completamente en un notable folleto de Dedekind cuya lectura es fascinante aún hoy en día.[1]

En efecto, Dedekind mostró que el tratamiento intuitivo e "ingenuo" seguido por todos los grandes matemáticos desde Fermat y Newton hasta Gauss y Riemann estaba sobre la pista correcta: que el sistema de números reales (como símbolos para las longitudes de segmentos, o definidos de alguna otra manera) es un instrumento coherente y completo para la medición científica, y que en este sistema las reglas de cálculo del sistema de números racionales siguen siendo válidas.

Sin perjuicio, podríamos dejarlo así y dirigirnos directamente a la esencia del cálculo. Sin embargo, para un entendimiento más profundo del concepto de número real, el cual es necesario para nuestro trabajo posterior, la siguiente explicación, así como el Suplemento a este capítulo, deberían ser estudiados.

b. Números reales e intervalos encajados

Por el momento imaginemos los puntos en una recta L como los elementos básicos del continuo. Postulamos el que a cada punto en L le corresponde un "número real" x, su *coordenada,* y que para estos números x, y las relaciones recién descritas para los números racionales conservan su significado. En particular, la relación $x < y$ indica orden en L y la expresión $|y - x|$ significa la distancia entre el punto x y el punto y. El problema básico es el de relacionar estos números (o mediciones en el continuo de puntos dado geométricamente) con los números racionales considerados originalmente y, por tanto, finalmente con los enteros. Además, debemos explicar cómo operar con los elementos de este "continuo de números" en la misma forma que con los números racionales. Con el tiempo, formularemos el concepto del continuo independientemente de los conceptos geométricos intuitivos, pero por ahora postergamos algo de la discusión más abstracta hasta el Suplemento.

¿Cómo podemos describir un número real irracional? Para algunos números, tales como $\sqrt{2}$ o π, podemos dar una caracterización geométrica simple, pero esto no es siempre factible. Un método suficientemente flexible para obtener cualquier punto real x consiste en describir el valor x por una sucesión de aproximaciones racionales de precisión cada vez mayor. Específicamente, aproximaremos x simultáneamente desde la derecha y desde la izquierda incrementando sucesivamente la precisión y de tal manera que el margen de error se aproxime a cero. En otras palabras, utilizamos una

[1] R. Dedekind, "Nature and Meaning of Number", en *Essays on Number,* Londres y Chicago, 1901. (El primero de estos ensayos, "Continuity and Irrational Numbers", proporciona una información detallada de la definición y leyes de operación con números reales.) Reimpreso bajo el título *Essays on the Theory of Numbers,* Dover, Nueva York, 1964. El original de estas traducciones apareció en 1887 bajo el título "Was sind und was sollen die Zahlen?"

"sucesión" de intervalos racionales que contienen a x, con cada intervalo de la sucesión conteniendo al siguiente de tal manera que la longitud del intervalo, y con él el error de aproximación, puede hacerse, tomando intervalos suficientemente alejados en la sucesión, menor que cualquier número positivo especificado.

Para comenzar, confínese x a un intervalo cerrado $I_1 = [a_1, b_1]$, esto es,

$$a_1 \leq x \leq b_1,$$

donde a_1 y b_1 son racionales (ver Fig. 1.3). Dentro de I_1 consideramos un "subintervalo" $I_2 = [a_2, b_2]$ que contiene a x, esto es,

$$a_1 \leq a_2 \leq x \leq b_2 \leq b_1,$$

donde a_2 y b_2 son racionales. Por ejemplo, podemos escoger para I_2 una de las mitades de I_1, pues x debe encontrarse en uno o ambos de los semi-intervalos. Dentro de I_2 consideramos un subintervalo $I_3 = [a_3, b_3]$ que también contiene a x:

$$a_1 \leq a_2 \leq a_3 \leq x \leq b_3 \leq b_2 \leq b_1,$$

donde a_3 y b_3 son racionales, etc. Se requiere que la longitud del intervalo I_n *tienda a cero* con n creciente; esto es, que la longitud de I_n sea menor que cualquier número positivo preasignado, para toda n suficientemente grande. Un conjunto de intervalos cerrados $I_1, I_2, I_3, \ldots$, cada uno conteniendo al siguiente y tal que las longitudes tiendan a cero se denominará "sucesión anidada de intervalos" o "encaje" de intervalos. El punto x está determinado de manera única por la sucesión o encaje; esto es, ningún otro



Figura 1.3 Una sucesión anidada de intervalos.

punto y puede pertenecer a todo I_n, puesto que la distancia entre x e y excede la longitud de I_n una vez que n es suficientemente grande. Puesto que aquí escogemos siempre puntos racionales para los puntos extremos de los I_n y puesto que todo intervalo con puntos extremos racionales es descrito por dos números racionales, vemos que todo punto x de L, esto es, todo número real, puede ser precisamente determinado con la ayuda de una infinidad de números racionales. La aseveración inversa no es tan obvia; la aceptaremos como un *axioma* básico.

POSTULADO DE LOS ENCAJES DE INTERVALOS. *Si $I_1, I_2, I_3, \ldots$ forman una sucesión "encaje" de intervalos con puntos extremos racionales, existe un punto x contenido en todo I_n.*[1]

[1] Es importante destacar que para un encaje los intervalos I_n son *cerrados*. Si, por ejemplo, I_n denota el intervalo abierto $0 < x < 1/n$, entonces cada I_n contiene al siguiente y las longitudes de los intervalos tienden a cero, pero no existe x contenido en todo I_n.

Como veremos, éste es un *axioma de continuidad:* garantiza que no existen vacíos en el eje real. Usaremos el axioma para caracterizar el continuo real y justificar todas las operaciones con límites que son básicas para el cálculo y el análisis. (Como veremos posteriormente, hay también muchas otras maneras de formular este axioma.)

c. Fracciones decimales. Bases distintas a la de diez

Fracciones decimales infinitas. Una de las muchas maneras de definir números reales es la descripción familiar en términos de decimales infinitos. Es enteramente posible tomar los decimales infinitos como los objetos básicos, en lugar de los puntos del eje numérico; pero procederíamos tal vez en una forma geométrica más sugestiva definiendo la representación decimal infinita de números reales en términos de encajes de intervalos.

Consideremos el eje numérico subdividido en intervalos unitarios por los puntos correspondientes a enteros. Un punto x se encuentra entre dos puntos sucesivos de subdivisión, o bien es él mismo uno de los puntos de división. En cualquier caso existe por lo menos un entero c_0 tal que

$$c_0 \leq x \leq c_0 + 1,$$

de modo que x pertenece al intervalo cerrado $I_0 = [c_0, c_0 + 1]$. Dividimos I_0 en diez partes iguales por puntos $c_0 + 1/10$, $c_0 + 2/10, \ldots, c_0 + 9/10$. El punto x debe entonces pertenecer por lo menos a uno de los subintervalos cerrados de I_0 (posiblemente a dos adyacentes, si x es uno de los puntos de subdivisión). En otras palabras, existe un dígito c_1 (esto es, uno de los enteros $0, 1, 2, \ldots, 9$) tal que x pertenece al intervalo cerrado I_1 dado por

$$c_0 + 1/10c_1 \leq x \leq c_0 + 1/10c_1 + 1/10.$$

Dividiendo a su vez I_1 en diez partes iguales encontramos un dígito c_2 tal que x se encuentra en el intervalo I_2 dado por

$$c_0 + 1/10c_1 + 1/100c_2 \leq x \leq c_0 + 1/10c_1 + 1/100c_2 + 1/100.$$

Repetimos este proceso. Después de n pasos x está confinado en un intervalo I_n dado por

$$c_0 + \frac{1}{10}c_1 + \cdots + \frac{1}{10^n}c_n \leq x \leq c_0 + \frac{1}{10}c_1 + \cdots + \frac{1}{10^n}c_n + \frac{1}{10^n},$$

donde $c_1, c_2, \ldots$ son todos dígitos. El intervalo I_n tiene longitud $1/10^n$, la cual tiende a cero para n creciente. Es claro que los I_n forman un encaje de intervalos y, por lo tanto, que x está determinado de manera única por los I_n. Puesto que los I_n son conocidos, una vez que los números $c_0, c_1, c_2, \ldots$ están dados, encontramos que un número real arbitrario puede ser descrito completamente por una sucesión infinita de enteros $c_0, c_1, c_2, \ldots$, donde

todos excepto el primero son dígitos que tienen valores desde cero hasta nueve solamente. En la notación decimal ordinaria, la conexión entre x y $c_0, c_1, c_2, \ldots$ se indica escribiendo

$$x = c_0 + 0.c_1c_2c_3 \cdots .$$

(Usualmente el entero c_0 mismo es también escrito en notación decimal si c_0 es positivo.) Inversamente, por el axioma de continuidad, toda aquella expresión que denota una fracción decimal infinita representa un número real.

Es posible que existan dos representaciones decimales diferentes del mismo número; por ejemplo,

$$1 = 0.99999 \cdots = 1.00000 \cdots .$$

En nuestra construcción el entero c_0 está determinado de manera única por x, a menos que el propio x sea un entero. En este caso podríamos elegir $c_0 = x$ o bien $c_0 = x - 1$. Una vez efectuada la elección, el dígito c_1 es único a menos que x sea uno de los nuevos puntos que subdividen I_0 en diez partes iguales. Continuando así, encontramos que c_0 y todos los c_k están determinados de manera única por x a menos que x resulte un punto de subdivisión en algún paso. Si esto sucediera por primera vez en el n-ésimo paso, entonces

$$x = c_0 + \frac{1}{10}c_1 + \cdots + \frac{1}{10^n}c_n,$$

donde $c_1, c_2, \ldots, c_n$ son dígitos y donde $c_n > 0$, puesto que de otra manera x hubiera sido un punto de subdivisión en algún paso anterior. Se sigue que I_{n+1} es el intervalo $[x, x + 1/10^{n+1}]$ o bien el intervalo $[x - 1/10^{n+1}, x]$. En el primer caso x será el punto extremo izquierdo de todos los intervalos posteriores $I_{n+2}, I_{n+3}, \ldots$, y en el segundo caso, el punto extremo derecho. Esto nos lleva entonces a la representación decimal

$$x = c_0 + 0.c_1c_2 \cdots c_n000 \cdots$$

o bien a la representación

$$x = c_0 + 0.c_1c_2 \cdots (c_n - 1)99999 \cdots .$$

De aquí que el único caso en el cual puede surgir una ambigüedad es para un número racional x que pueda ser escrito como una fracción que tiene una potencia de diez para su denominador. Podemos aún eliminar esta ambigüedad excluyendo representaciones decimales en las cuales todos los dígitos a partir de cierto punto son nueves.

En la representación decimal de números reales el papel especial desempeñado por el número diez es puramente incidental. La única razón evidente para el extenso uso del sistema decimal es la facilidad de contar por

decenas con nuestros dedos (dígitos). Todo entero p mayor que uno puede servir igualmente bien. Podrían utilizarse p subdivisiones iguales en cada paso. Un número real x sería entonces representado en la forma

$$x = c_0 + 0.c_1c_2c_3\cdots,$$

donde c_0 es un entero, y ahora $c_1, c_2, \ldots$ tienen uno de los valores 0, 1, 2, ..., $p - 1$. Esta representación caracteriza nuevamente a x por un encaje de intervalos, a saber

$$c_0 + \frac{1}{p}c_1 + \cdots + \frac{1}{p^n}c_n \leq x \leq c_0 + \frac{1}{p}c_1 + \cdots + \frac{1}{p^n}c_n + \frac{1}{p^n}.$$

Si x es positivo o cero, el entero c_0 es también positivo o cero, y el propio c_0 tiene un desarrollo finito de la forma

$$c_0 = d_0 + pd_1 + p^2d_2 + \cdots + p^kd_k,$$

donde $d_0, d_1, \ldots, d_k$ toman uno de los valores $0, 1, \ldots, p - 1$. La representación completa de x "en la base p" toma la forma

$$x = d_kd_{k-1}\cdots d_1d_0.c_1c_2c_3\cdots.$$

Si x es negativo podemos utilizar este tipo de representación para $-x$.

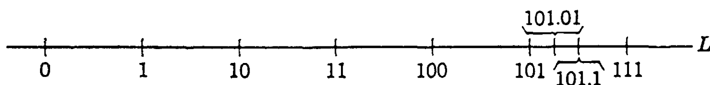


Figura 1.4 La fracción ²¹⁄₄ en el sistema binario.

Bases distintas a la de 10 han sido utilizadas extensamente en la actualidad. Siguiendo la dirección de los antiguos babilonios, los astrónomos durante muchos siglos representaron sistemáticamente los números como fracciones "sexagesimales", con $p = 60$ como la base.

Representación binaria. El sistema "binario" con la base $p = 2$ tiene un especial interés teórico y es útil en el diseño lógico de máquinas calculadoras electrónicas. En el sistema binario los dígitos tienen solamente dos valores posibles, el cero y el uno. El número ²¹⁄₄, por ejemplo, sería escrito 101.01, de acuerdo con la fórmula

$$\frac{21}{4} = 2^2\cdot 1 + 2^1\cdot 0 + 1\cdot 1 + \frac{1}{2}\cdot 0 + \frac{1}{2^2}\cdot 1 \qquad \text{(ver Fig. 1.4)}.$$

Cálculo con números reales. Aunque la definición de números reales y sus representaciones decimal infinita, binaria, etc., son naturales, puede no parecer obvio el que uno pueda operar con el continuo de números exactamente como con números racionales, efectuando las operaciones racionales

y reteniendo las leyes de la aritmética, tales como las leyes asociativa, conmutativa y distributiva. La demostración es simple aunque algo tediosa. En lugar de dificultar el camino hacia la esencia viva del análisis considerando la cuestión aquí, aceptaremos temporalmente la posibilidad de cálculo aritmético ordinario con los números reales. Una comprensión más profunda de la estructura lógica subyacente al concepto de número será alcanzada cuando se descubra la idea de límite y sus implicaciones. (Ver el Suplemento a este capítulo, p. 112.)

d. Definición de vecindad o entorno

No sólo las operaciones racionales, sino también relaciones de orden o desigualdades para números reales obedecen las mismas reglas que los números racionales.

Pares de números reales a y b con $a < b$ dan nuevamente lugar a intervalos cerrados $[a, b]$ (dados por $a \leq x \leq b$) e intervalos abiertos (a, b) (dados por $a < x < b$). Frecuentemente tendremos necesidad de asociar con un punto x_0 los diferentes intervalos abiertos que contienen a ese punto o bien, específicamente, tenerlo como centro; a esos intervalos los llamaremos *vecindades* o *entornos* del punto. Más precisamente, para todo positivo ε, la *ε-vecindad* del punto x_0 consiste de los valores x para los cuales $x_0 - \varepsilon < x < x_0 + \varepsilon$; esto es, es el intervalo $(x_0 - \varepsilon, x_0 + \varepsilon)$. Todo intervalo abierto (a, b) que contiene a un punto x_0, también contiene a una vecindad completa de x_0.

Habiendo definido intervalos con puntos extremos reales, podemos formar ahora *encajes* de intervalos utilizando la misma definición como en el caso de puntos extremos racionales. Es de lo más importante para la coherencia lógica del cálculo que para todo encaje de intervalos con puntos extremos reales exista un número real contenido en todos ellos. (Ver Suplemento, p. 118.)

e. Desigualdades

Reglas básicas

Las desigualdades juegan un papel mucho mayor en matemáticas superiores que en matemáticas elementales. A menudo el valor preciso de una cantidad x es difícil de determinar, mientras que puede ser fácil efectuar una estimación de x, esto es, mostrar que x es mayor que alguna cantidad conocida a y menor que alguna otra cantidad b. Para muchos propósitos sólo la información contenida en tal estimación de x es significativa. Recordaremos por ello brevemente algunas de las reglas elementales acerca de las desigualdades.

El hecho básico es que la suma y producto de dos números reales positivos son también números positivos,* esto es, si $a > 0$ y $b > 0$, entonces $a + b > 0$ y $ab > 0$. Más aún, contamos con el hecho de que la desigualdad $a > b$ es equivalente con $a - b > 0$. Consecuentemente, dos desigualdades $a > b$ y $c > d$ pueden ser *sumadas* para dar la desigualdad $a + c > b + d$, puesto que

$$(a + c) - (b + d) = (a - b) + (c - d)$$

es positivo por ser la suma de dos números positivos. (Restar las desigualdades para obtener $a - c > b - d$ no es legítimo. ¿Por qué?) Una desigualdad puede ser *multiplicada por un número positivo;* esto es, si $a > b$ y $c > 0$, entonces $ac > bc$. Para la demostración, observamos que

$$ac - bc = (a - b)c$$

es positivo puesto que es el producto de números positivos. Si c es *negativo,* podemos concluir de $a > b$ que $ac < bc$. Más generalmente, se sigue de $a > b > 0$ y $c > d > 0$ que $ac > bd$.

Es obvio geométricamente que la desigualdad es *transitiva;* esto es, si $a > b$ y $b > c$, entonces $a > c$. La transitividad[1] se sigue también inmediatamente a partir de la positividad de la suma

$$(a - b) + (b - c) = a - c.$$

Las reglas anteriores también se cumplen si se reemplaza el signo $>$ por $\geq$ en todas partes.

Sean a y b números *positivos* y obsérvese que

$$a^2 - b^2 = (a + b)(a - b).$$

Puesto que $a + b$ es positivo, concluimos que $a^2 > b^2$ es consecuencia de tener $a > b$. Por tanto, una desigualdad entre números positivos puede ser "elevada al cuadrado". Análogamente, $a^2 \geq b^2$ siempre que $a \geq b \geq 0$. De la ecuación

$$a - b = \frac{1}{a + b}(a^2 - b^2),$$

válida para todos a y b positivos, se sigue que la recíproca también es cierta; esto es, para a y b positivos, $a^2 > b^2$ implica $a > b$. Aplicando este resultado a los números $a = \sqrt{x}$, $b = \sqrt{y}$, para números positivos arbitrarios x, y,

---

* (N. del T.) El conjunto de los números positivos, con las operaciones $+$ y $\cdot$, constituye un *anillo,* que es un *subanillo* de los reales. Para estos conceptos véase cualquier libro de introducción al álgebra moderna.

[1] La transitividad justifica el uso de la fórmula compuesta "$a < b < c$" para expresar "$a < b$ y $b < c$, etc.". Evítense arreglos no transitivos como $x < y > z$; éstos dan lugar a confusiones y engaños.

encontramos[1] que $\sqrt{x} > \sqrt{y}$ cuando $x > y$. Más generalmente, $\sqrt{x} \geq \sqrt{y}$ siempre que $x \geq y \geq 0$. De aquí que es legítimo tomar raíz cuadrada de ambos miembros de una desigualdad entre números reales no negativos.

Supóngase que a y b son positivos y n es un entero positivo. En la factorización

$$a^n - b^n = (a - b)(a^{n-1} + a^{n-2}b + \cdots + b^{n-1})$$

el segundo factor es positivo. Por ello $a^n - b^n$ tiene el mismo signo que $a - b$; si $a^n > b^n$, entonces $a > b$ y si $a^n < b^n$, entonces $a < b$.

Muchas desigualdades que encontraremos aparecen en la forma de estimaciones para el valor absoluto de un número. Recordemos que $|x|$ se define como x para $x \geq 0$ y $-x$ para $x < 0$. Puede decirse también que $|x|$ es el mayor de los números x y $-x$ cuando x no es cero y es igual a ambos cuando x es cero. La desigualdad $|x| \leq a$ establece entonces que ni x ni

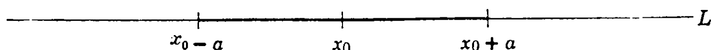


Figura 1.5 El intervalo $|x - x_0| \leq a$.

$-x$ excede a, esto es, que $x \leq a$ y $-x \leq a$. Puesto que $-x \leq a$ es equivalente a $x \geq -a$, vemos que la desigualdad $|x| \leq a$ significa que x se encuentra en el intervalo cerrado $-a \leq x \leq a$ con centro 0 y longitud $2a$. La desigualdad $|x - x_0| \leq a$ establece entonces que $-a \leq x - x_0 \leq a$, o bien que $x_0 - a \leq x \leq x_0 + a$; por lo tanto, que x se encuentra en el intervalo cerrado con centro x_0 y longitud $2a$ (ver Fig. 1.5). Análogamente, la ε-vecindad $(x_0 - \varepsilon, x_0 + \varepsilon)$ de un punto x_0, esto es, el intervalo abierto $x_0 - \varepsilon < x < x_0 + \varepsilon$, puede ser descrita por la desigualdad $|x - x_0| < \varepsilon$.

Desigualdad del triángulo

Una de las desigualdades más importantes que involucra valores absolutos es la llamada desigualdad del triángulo

$$|a + b| \leq |a| + |b|$$

para a, b, reales cualesquiera. El nombre "desigualdad del triángulo" es más apropiado para el aserto equivalente

$$|\alpha - \beta| \leq |\alpha - \gamma| + |\gamma - \beta|$$

para el cual hemos hecho $a = \alpha - \gamma$, $b = \gamma - \beta$. La interpretación geométrica de esta desigualdad es que la distancia directa desde α hasta β es menor

[1] Aquí y en lo sucesivo el símbolo $\sqrt{z}$ para $z \geq 0$ denota aquel número no negativo cuyo cuadrado es z. Con esta convención $|c| = \sqrt{c^2}$ para todo real c puesto que $|c| \geq 0$ y $|c|^2 = c^2$. De esto se obtiene la importante identidad $|xy| = |x| \cdot |y|$ puesto que

$$|xy|^2 = (xy)^2 = x^2y^2 = (|x| \cdot |y|)^2.$$

que o igual a la suma de las distancias vía un tercer punto γ; (esto corresponde también al hecho de que en cualquier triángulo la suma de dos de sus lados excede al tercero).

Una demostración formal de la desigualdad del triángulo es fácil de dar. Distinguimos los casos $a + b \geq 0$ y $a + b < 0$. En el primer caso la desigualdad establece que $a + b \leq |a| + |b|$; pero esto se obtiene trivialmente mediante la suma de las desigualdades $a \leq |a|$ y $b \leq |b|$. En el segundo caso la desigualdad del triángulo se reduce a $-(a + b) \leq |a| + |b|$, la cual se obtiene nuevamente mediante la suma de $-a \leq |a|$, $-b \leq |b|$.

Obtenemos inmediatamente una desigualdad análoga para tres cantidades:

$$|a + b + c| \leq |a| + |b| + |c|;$$

pues, aplicando la desigualdad del triángulo dos veces, resulta

$$|a + b + c| = |(a + b) + c| \leq |a + b| + |c| \leq |a| + |b| + |c|.$$

De la misma manera se obtiene la desigualdad más general

$$|a_1 + a_2 + \cdots + a_n| \leq |a_1| + |a_2| + \cdots + |a_n|$$

Ocasionalmente se requieren estimaciones para $|a + b|$ por debajo. Observamos que

$$|a| = |(a + b) + (-b)| \leq |a + b| + |-b| = |a + b| + |b|,$$

y por lo tanto, que la desigualdad

$$|a + b| \geq |a| - |b|$$

se satisface.

La desigualdad de Cauchy-Schwarz

Algunas de las desigualdades más importantes utilizan el hecho obvio de que el cuadrado de un número real nunca es negativo y que, consecuentemente, una suma de cuadrados tampoco puede ser negativa. Uno de los resultados más frecuentemente utilizados, obtenido de esta manera, es la desigualdad de Cauchy-Schwarz:

$$(a_1b_1 + a_2b_2 + \cdots + a_nb_n)^2 \leq (a_1^2 + a_2^2 + \cdots + a_n^2)(b_1^2 + b_2^2 + \cdots + b_n^2).$$

Haciendo

$$A = a_1^2 + a_2^2 + \cdots + a_n^2,$$
$$B = a_1b_1 + a_2b_2 + \cdots + a_nb_n,$$
$$C = b_1^2 + b_2^2 + \cdots + b_n^2,$$

la desigualdad se transforma en $AC \geq B^2$. Para probarla observamos que para todo t real es

$$0 \leq (a_1 + tb_1)^2 + (a_2 + tb_2)^2 + \cdots + (a_n + tb_n)^2$$

puesto que el miembro derecho es una suma de cuadrados. Desarrollando cada cuadrado y ordenando según las potencias de t, se encuentra que

$$0 \leq A + 2Bt + Ct^2$$

para todo t, donde A, B, C tienen el mismo significado anterior. Aquí es $C \geq 0$. Podemos suponer que $C > 0$, pues, ciertamente, $B^2 = AC = 0$ cuando $C = 0$. Substituyendo entonces para t el valor especial $t = -B/C$ [que corresponde al mínimo de la expresión cuadrática

$$A + 2Bt + Ct^2 = C\left[\left(t + \frac{B}{C}\right)^2 + \left(A - \frac{B^2}{C}\right)\right],$$

encontramos que

$$0 \leq A - \frac{2B^2}{C} + \frac{B^2}{C} = \frac{AC - B^2}{C},$$

y por lo tanto $AC - B^2 \geq 0$.

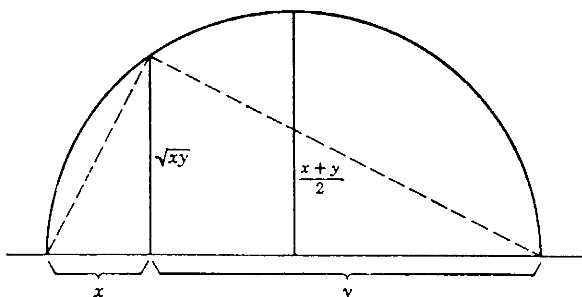


Figura 1.6 Medias geométrica y aritmética de x e y.

En el caso especial $n = 2$ podemos escoger

$$a_1 = \sqrt{x}, \qquad a_2 = \sqrt{y}, \qquad b_1 = \sqrt{y}, \qquad b_2 = \sqrt{x},$$

donde x e y son números positivos. La desigualdad adquiere entonces la forma $(2\sqrt{xy})^2 \leq (x + y)^2$, o bien

$$\sqrt{xy} \leq \frac{x + y}{2}.$$

Esta desigualdad establece que la *media geométrica* $\sqrt{xy}$ de dos números positivos, x, y, no excede nunca su *media aritmética* $(x + y)/2$. La media geométrica de dos números, x, y, puede ser interpretada como la longitud de la altura en un triángulo rectángulo, cuya proyección sobre la hipotenusa la divide en segmentos de longitud x e y respectivamente. Esta desigualdad

establece entonces que en un triángulo rectángulo la altura no excede la mitad de la hipotenusa (ver Fig. 1.6).[1]

1.2 El concepto de función

Desde el comienzo de las matemáticas modernas en el siglo XVII el concepto de función ha estado en el centro mismo del pensamiento matemático. (Leibnitz parece haber sido el primero en utilizar la palabra "función".) Aunque la idea de las relaciones funcionales es significativa mucho más allá del dominio matemático, enfocaremos naturalmente nuestra atención sobre funciones en el sentido matemático, esto es, sobre la conexión entre cantidades matemáticas por relaciones matemáticas o prescripciones u "operaciones". Una gran parte de las matemáticas y las ciencias naturales es dominada por relaciones funcionales, pues aparecen en todo el Análisis, la Geometría, la Mecánica y otros campos. Por ejemplo, la presión en un gas ideal es una función de la densidad y de la temperatura; la posición de una molécula móvil es una función del tiempo; el volumen y la superficie de un cilindro son funciones de su radio y de su altura. Siempre que los valores de ciertas cantidades, $a, b, c, \ldots$, estén determinados por los de ciertas otras, $x, y, z, \ldots$, decimos que $a, b, c, \ldots$ *dependen* de $x, y, z, \ldots$, o bien que son *funciones* de $x, y, z, \ldots$. Ejemplos de relaciones funcionales están dados mediante expresiones formales tales como las siguientes:

a) La fórmula $A = a^2$ define A como función de a. Para $a > 0$, puede interpretarse A como el área de un cuadrado de lado a.

b) La fórmula

$$y = \sqrt{1 - x^2}$$

define y como función de x para todo x tal que $-1 \leq x \leq 1$. Para $x > 0$ esta función expresa el lado y de un triángulo rectángulo con hipotenusa 1, en términos del otro lado x.

c) Las ecuaciones

$$x = t, \qquad y = -t^2$$

asignan valores a x y a y para cada t, y definen por ello a x y a y como funciones de t. Si se interpretan x e y como las coordenadas rectangulares de un punto P en el plano, y t como el tiempo, entonces nuestras ecuaciones describen la localización de P en el tiempo t; en otras palabras, describen el *movimiento* del punto P.

d) Las ecuaciones

$$a = \frac{x}{x^2 + y^2}, \qquad b = \frac{y}{x^2 + y^2}$$

[1] El lector interesado encontrará más material en *An Introduction to Inequalities,* por E. F. Beckenbach y R. Bellman, Random House, 1961, y en *Geometric Inequalities,* por N. Kazarinoff, Random House, 1961.

definen a y b como funciones de x e y para $x^2 + y^2 \neq 0$. Interpretando las parejas de valores (x, y) y (a, b) como coordenadas rectangulares de dos puntos, vemos que las ecuaciones asignan a cada punto (x, y) [con excepción del origen $(0, 0)$] una "imagen" (a, b). El lector puede verificar fácilmente que la imagen (a, b) se encuentra siempre en el mismo rayo desde el origen que el "original" o "antecedente" (x, y) y tiene la distancia recíproca desde el origen. Se habla de *representar* o "mapear" (x, y) sobre (a, b) por medio de las ecuaciones que expresan a y b en términos de x e y.

En los ejemplos anteriores la ley funcional es expresada por fórmulas sencillas que determinan ciertas cantidades en términos de otras.[1] Las cantidades que aparecen en los miembros izquierdos, "las variables dependientes", son expresadas en términos de las "variables independientes", en el miembro derecho. La ley matemática que asigna valores únicos de las variables dependientes a valores dados de las variables independientes se llama una *función.* Esta no es afectada por los nombres x, y, etc., dados a estas variables. En el ejemplo *c)* se tiene una variable independiente t y dos variables dependientes x, y, mientras que en el ejemplo *d)* hay dos variables independientes, x e y y dos variables dependientes, a y b.

La dependencia de y con respecto de x, por una relación funcional, es indicada frecuentemente por la expresión breve: "y es función de x".[2]

a. Función-Gráfica

Dominio y rango de una función

Las variables independientes usualmente se interpretan geométricamente como coordenadas de un punto en una o más dimensiones. En el ejemplo *b)* éste sería un punto en el eje x; en el ejemplo *d)*, un punto en el plano x, y. En ocasiones, las variables independientes son libres de tomar todos los valores, como en los ejemplos *a)* y *c)*. Sin embargo, a menudo existe alguna restricción, inherente, o bien impuesta, y las funciones no están definidas para todos los valores. El conjunto de valores o los puntos para los cuales una función está definida constituye el "*dominio*" de la función. En el ejemplo *a)* el dominio es todo el eje a; en el *b)*, el intervalo $-1 \leq x \leq 1$; en el *c)*, todo el eje t; y en el ejemplo *d)*, los puntos del plano x, y distintos del origen.

---

[1] Más adelante se comprenderá gradualmente la necesidad de considerar funciones que no son susceptibles de tales representaciones mediante fórmulas sencillas. (Véase, por ejemplo, p. 49.)

[2] Esta locución es utilizada libremente en las ciencias, pero algunos de los textos más pedantes la evitan. No tiene objeto estorbarnos a nosotros mismos por un desmedido afán de tener una "precisión" quisquillosa cuando ésta no tiene relación con la esencia del asunto.

A cada punto P en el dominio las funciones asignan valores definidos para las variables dependientes. Estos valores pueden ser también interpretados como coordenadas de un punto Q, la *imagen* de P. Se dice que P es representado o transformado o "mapeado" por las funciones sobre el punto Q. Por lo tanto, en el ejemplo *d)* el punto $P = (1, 2)$ del plano x, y es transformado en el punto $Q = (1/5, 2/5)$ del plano a, b. Los puntos imagen Q constituyen el *rango* de la función.[1] Cada Q en el rango es la imagen de uno (o más) puntos en el dominio de la función.

En el ejemplo *c)*, puntos del eje t tienen como imágenes puntos en el plano x, y. El eje t es "mapeado" *en* el plano x, y. Pero no todo punto en el plano x, y aparece como imagen, sino sólo aquéllos para los cuales $y = -x^2$. Por lo tanto, el rango de la función es la parábola $y = -x^2$. Se dice que el eje t es transformado o "mapeado" *sobre* la parábola $y = -x^2$, en el sentido de que los puntos imagen llenan esta parábola.

En el ejemplo *d)*, el rango consiste de los puntos (a, b) en el plano a, b cuyas coordenadas pueden escribirse en la forma $a = x/(x^2 + y^2)$, $b = y/(x^2 + y^2)$, con x, y escogidos de tal manera que $x^2 + y^2 \neq 0$. En otras palabras, el rango consiste de aquellos puntos (a, b) para los cuales las ecuaciones anteriores tienen una solución (x, y). Como se ve inmediatamente, el rango consiste de los puntos (a, b) para los cuales a y b no se hacen cero simultáneamente; cada uno de tales puntos (a, b) es imagen del punto $x = a/(a^2 + b^2)$, $y = b/(a^2 + b^2)$. Toda figura geométrica en el plano x, y es transformada entonces sobre una figura correspondiente en el plano a, b, la cual consiste de las imágenes de los puntos de la primera figura. Por ejemplo, un círculo $x^2 + y^2 = r^2$ en torno al origen es transformado en el círculo $a^2 + b^2 = 1/r^2$ en el plano a, b.

En éste y en los siguientes capítulos trataremos casi exclusivamente con una sola variable independiente, digamos, x, y una sola variable dependiente, digamos, y, como se indica en el ejemplo *b)*.[2] Comúnmente, tal función se representa en la forma usual por su *gráfica* en el plano x, y, esto es, por la curva que consiste de aquellos puntos (x, y) cuya ordenada se encuentra en la relación funcional especificada con la abscisa x (ver Fig. 1.7). Para el ejemplo *b)* la gráfica es la mitad superior de un círculo de radio uno en torno al origen.

La interpretación de función como transformación o representación de un dominio en el eje x sobre un rango en el eje y conduce a una visualización diferente de las funciones. Interpretamos a x e y no como coordenadas del mismo punto en el plano x, y, sino como puntos en dos diferentes

---

[1] Es conveniente a menudo hablar del punto Q como de "una función" de P, no obstante que en la representación analítica aparecen varias funciones que expresan las diferentes coordenadas de Q.

[2] Sin embargo, se deberá recalcar desde el principio que funciones de varias variables aparecen de la misma natural manera en muchos casos. Estas serán discutidas sistemáticamente en el Volumen II.

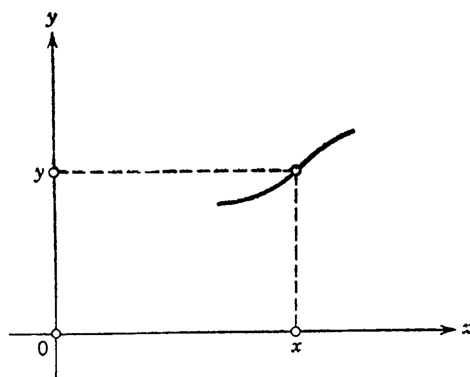


Figura 1.7 Gráfica de una función.

ejes numéricos independientes. Entonces la función transforma o representa un punto x del eje x en un punto y sobre el eje y. Tales transformaciones surgen frecuentemente en la geometría, tal como la transformación "afín", la cual se produce proyectando un punto x del eje x sobre un punto y en un eje y paralelo, desde un centro 0 localizado en el plano de los dos ejes (ver Fig. 1.8). Esta transformación puede ser expresada analíticamente, como es fácil comprobar, por la función lineal $y = ax + b$ con a y b cons-

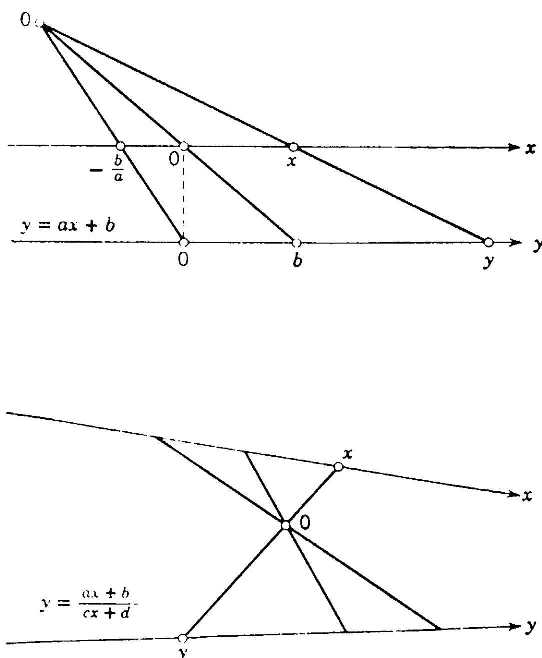


Figura 1.8 Mapeos.

tantes. Obviamente, es una función *biunívoca* o "uno a uno", en la cual inversamente al punto imagen y le corresponde un único punto x original.* Otro ejemplo más general es la "transformación *perspectiva*" definida por el mismo tipo de proyección, salvo que los dos ejes no son necesariamente paralelos. En este caso la expresión analítica está dada por la función racional lineal de la forma $y = (ax + b)/(cx + d)$, con a, b, c y d constantes.

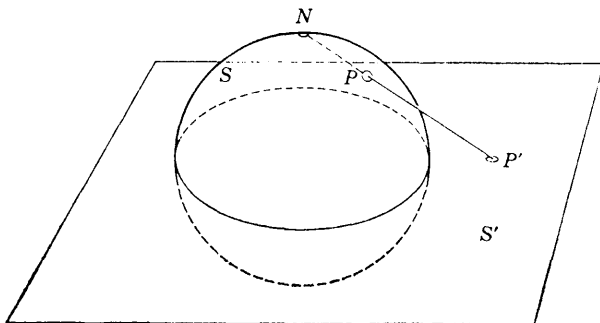


Figura 1.9 Proyección estereográfica.

Cualquier proyección de una superficie S en el espacio sobre otra superficie, S', desde algún centro N, puede ser vista como una transformación o función cuyo dominio es S y cuyo rango se encuentra en S'. Por ejemplo, podemos transformar o "representar" una esfera sobre un plano ecuatorial, proyectando cada punto P de la esfera sobre un punto P' del plano mediante rayos desde el Polo Norte (ver Fig. 1.9). Esta representación o "mapeo" es la "proyección estereográfica", que se utiliza frecuentemente para mapas de la Tierra. Ejemplos de este tipo sugieren la interpretación de funciones como "mapeos" o transformaciones.

Cuando intervienen más variables independientes o dependientes, la definición de funciones por transformaciones o "mapeos" proporciona una interpretación más flexible y adecuada que aquélla por gráficas. Este hecho resultará totalmente claro en el segundo volumen.

b. Definición del concepto de funciones de una variable continua. Dominio y rango de una función

Una función de una sola variable independiente x asigna valores y a valores x. El *dominio* de la función es la totalidad de valores x para los cuales la función está definida. En los casos que más nos interesan, el dominio de la función consiste de uno o varios intervalos (ver Fig. 1.10). Se

* (N. del T.) O sea, tanto la transformación dada como su *inversa* son "*unívocas*" (o univaluadas).

dice entonces que y es función de una *variable continua* (en contraste con otros casos donde, por ejemplo, la función puede estar definida sólo para valores racionales o enteros de x). Aquí, los "intervalos" que forman el dominio pueden o no contener sus puntos extremos, y pueden extenderse también al infinito en una o ambas direcciones.[1] Así, la función $y = \sqrt{1 - x^2}$ está definida en el intervalo cerrado $-1 \leq x \leq +1$; la función $y = 1/x$, en los dos intervalos abiertos semiinfinitos $x < 0$ y $x > 0$; la función $y = x^2$,

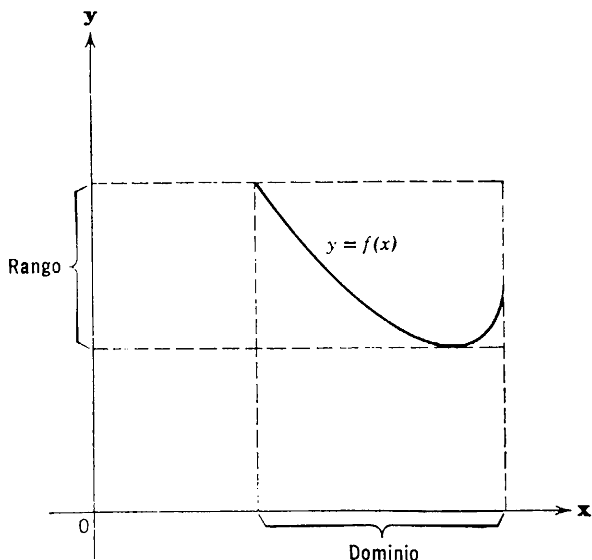


Figura 1.10 Dominio y rango de una función en representación gráfica.

en el 'intervalo" infinito $-\infty < x < +\infty$ que consiste de todo x; la función $y = \sqrt{(x^2 - 1)(4 - x^2)}$, en los dos intervalos separados $1 \leq x \leq 2$ y $-2 \leq x \leq -1$.

Las funciones se denotan por símbolos tales como f, F, g, etc. Las correspondientes relaciones entre x y los valores asociados y se escriben en la forma $y = f(x)$, o $y = F(x)$ o $y = g(x)$, etc., o también, en ocasiones, $y = y(x)$, para indicar[2] que y depende de x. Si, por ejemplo, $f(x)$ está definida por la expresión $x^2 + 1$, se tiene $f(3) = 3^2 + 1 = 10$, $f(-1) = (-1)^2 + 1 = 2$.

---

[1] Ordinariamente se reservará la palabra "intervalo" para el "acotado", esto es, para intervalos "finitos", que tienen puntos extremos que son números finitos definidos; entonces uno puede indicar el concepto más amplio, tal como se utiliza en el texto, por las palabras "conjuntos convexos", significando conjuntos que cuando contienen dos puntos deben contener todos los intermedios.

[2] En esta notación se trata de destacar las variables y no de indicar explícitamente la operación funcional por un símbolo tal como f. La notación

$$f: \quad x \to y$$

para la función f que mapea x en y se encuentra también ocasionalmente.

Naturaleza de la relación funcional

En la definición general de función, $f(x)$, no se dice nada respecto a la naturaleza de la relación mediante la cual se encuentra la variable dependiente cuando es dada la variable independiente. Como se dijo anteriormente, la función está dada a menudo en "forma cerrada" o explícita por una expresión sencilla como $f(x) = x^2 + 1$ o $f(x) = \sqrt{1 + \text{sen}^2 x}$, y en los primeros días del cálculo tales expresiones explícitas fueron casi exclusivamente lo que los matemáticos entendieron por funciones. A menudo, los dispositivos mecánicos generan curvas geométricas o bien gráficas que entonces definen funciones. Un ejemplo notable es la cicloide, una curva descrita por un punto fijo sobre un círculo que rueda a lo largo del eje x (ver Fig. 1.11). Su expresión analítica funcional por fórmulas se dará posteriormente (ver Sec. 4.1-*c*).

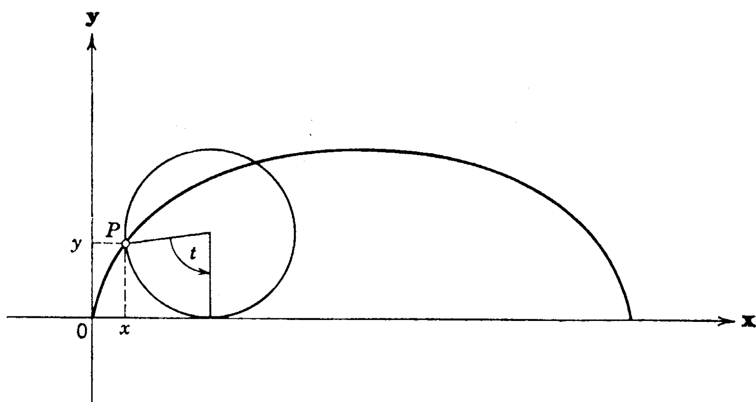


Figura 1.11.

Es claro que no estamos restringidos a tales funciones generadas geométrica o mecánicamente. *Cualquier regla o ley* mediante la cual un valor de y es asignado a valores de x constituye una función. En algunas investigaciones teóricas la amplia generalidad o bien la vaguedad del concepto de función es, de hecho, una ventaja. Sin embargo, para las aplicaciones, particularmente en el cálculo, el concepto general de función es innecesariamente amplio. Para hacer posibles desarrollos matemáticos significativos, las leyes "arbitrarias" de correspondencia mediante las cuales un valor de y es asignado a x deben estar sujetas a restricciones radicales. Durante el siglo pasado los matemáticos reconocieron y formularon en términos precisos las restricciones esenciales que había que imponer sobre tal concepto tan general para obtener funciones que verdaderamente tuvieran las propiedades útiles que uno intuitivamente podría esperar.

**Dominios de funciones extendidos o restringidos*

Aún para funciones dadas por fórmulas explícitas es importante darse cuenta de que cualquier descripción completa de una función debe incluir una definición del *dominio* de la función. Para nosotros la "función" f descrita por "$f(x) = x^2$ para $0 < x < 2$" no es estrictamente la misma función que la función g dada por "$g(x) = x^2$, en el dominio mayor $-2 < x < 2$", aun cuando $f(x)$ y $g(x)$ tengan los mismos valores en el intervalo $0 < x < 2$ donde ambas están definidas. Generalmente, una función f se llama una "restricción" de una función g (o bien g una "extensión" de f) si dondequiera que f está definida, g también está definida y toma los mismos valores. Por supuesto, la misma función f puede provenir por restricción de muchas funciones diferentes. En nuestro ejemplo anterior f es también una restricción de la función h definida por $h(x) = x^2$ para $0 < x < 2$, $h(x) = -x^2$ para $-2 < x \leq 0$. En efecto, este ejemplo ilustra el proceso inverso al de formar restricciones de una función, el cual puede denominarse "unión de secciones". Se pueden generar nuevas funciones definiéndolas simplemente por expresiones explícitas diferentes en distintas partes del dominio.

c. Representación gráfica. Funciones monótonas

La idea fundamental de la geometría analítica es dar una representación analítica a una curva definida originalmente por alguna propiedad geométrica. Esto se hace usualmente considerando una de las coordenadas rectangulares, digamos y, como una función $y = f(x)$ de la otra coordenada, x; por ejemplo, una parábola se representa por la función $y = x^2$; el círculo con radio 1 en torno del origen, por las dos funciones $y = \sqrt{1 - x^2}$ y $y = -\sqrt{1 - x^2}$. En el primer ejemplo puede pensarse de la función como definida en el intervalo infinito $-\infty < x < \infty$; en el segundo debemos restringirnos al intervalo $-1 \leq x \leq 1$, puesto que fuera de este intervalo la función no tiene sentido.[1]

Inversamente, si, en lugar de comenzar con una curva definida geométricamente, consideramos una función $y = f(x)$ dada analíticamente, podemos representar la dependencia funcional de y respecto de x gráficamente utilizando un sistema coordenado rectangular en la manera usual (ver Fig. 1.7). Si para cada abscisa x se toma la ordenada correspondiente $y = f(x)$, se obtiene la representación geométrica de la función. Las restricciones que han de imponerse al concepto de función deberán asegurar para su representación geométrica la forma de una curva geométrica "razonable". Esto, ciertamente, expresa una concepción intuitiva en lugar de una condición matemática estricta. Sin embargo, pronto se formularán condiciones, tales como continuidad, diferenciabilidad, etc., las cuales aseguran que la gráfica de una función es una curva capaz de ser visualizada geométricamente. Este no sería el caso si admitiésemos funciones "patológicas" tales como la siguiente: para todo valor racional de x la función y tiene el valor 1, y para

[1] Ordinariamente no se consideran valores imaginarios o complejos de x e y.

todo valor irracional de x el valor de y es 0.* Esta prescripción funcional asigna un valor definido de y a cada x; pero en todo intervalo de x, no importa cuán pequeño, el valor de y da saltos de 0 a 1 y de 1 a 0 una infinidad de veces. Este ejemplo demuestra que el concepto general no restringido de función da lugar a gráficas que no podríamos considerar como curvas.

Funciones multiformes

Consideraremos solamente funciones $y = f(x)$ que asignan un valor único de y a cada valor de x en el dominio, como, por ejemplo, $y = x^2$, o bien $y = \operatorname{sen} x$. Sin embargo, para una curva descrita geométricamente puede suceder lo que para el círculo $x^2 + y^2 = 1$: que una descripción completa de la curva no esté dada precisamente por una sola función (uniforme), sino que requiera varias funciones; para el caso del círculo, las dos funciones $y = \sqrt{1 - x^2}$ y $y = -\sqrt{1 - x^2}$. Lo mismo es cierto para la hipérbola $y^2 - x^2 = 1$, la cual está representada por las dos funciones $y = \sqrt{1 + x^2}$ e $y = -\sqrt{1 + x^2}$. Por eso, tales curvas no determinan sin ambigüedad las correspondientes funciones. Con frecuencia se dice que la curva está representada por una función *multiforme;* las funciones separadas que la representan se denominan entonces *ramas uniformes* de la función multiforme que representa la curva. Para mayor claridad se utilizará siempre la palabra "función" para indicar una función univaluada. Por ejemplo, el símbolo $\sqrt{x}$ (para $x \geq 0$) denotará siempre el número *no negativo* cuyo cuadrado es x.

Si una curva es la gráfica de *una* función, ésta es intersectada por cualquier paralela al eje y en a lo más un punto, puesto que a cada punto x en el intervalo de definición le corresponde justamente un valor de y. El círculo unitario representado por las dos funciones

$$y = \sqrt{1 - x^2} \qquad \text{e} \qquad y = -\sqrt{1 - x^2},$$

es intersectado por tales paralelas al eje y en más de un punto. Las porciones de una curva correspondientes a diferentes ramas uniformes son conectadas en ocasiones una con otra, de manera que la curva completa es una sola figura que puede ser dibujada con un trazo de la pluma; por ejemplo, el círculo (ver Fig. 1.12). Por otra parte, estas porciones pueden ser separadas completamente, como para la hipérbola (ver Fig. 1.13).

Ejemplos. Consideremos algunos ejemplos adicionales sobre la representación gráfica de funciones.

---

* (N. del T.) Esta es la llamada "función de Dirichlet".

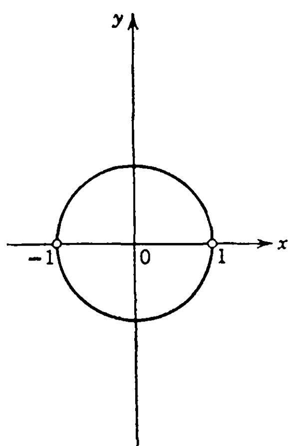


Figura 1.12.

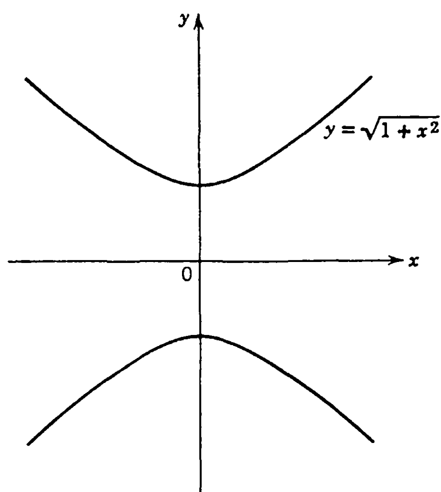


Figura 1.13.

a) y es proporcional a x,

$$y = ax.$$

La gráfica (ver Fig. 1.14) es una línea recta que pasa por el origen del sistema coordenado.

b) y es una "función lineal" de x,

$$y = ax + b.$$

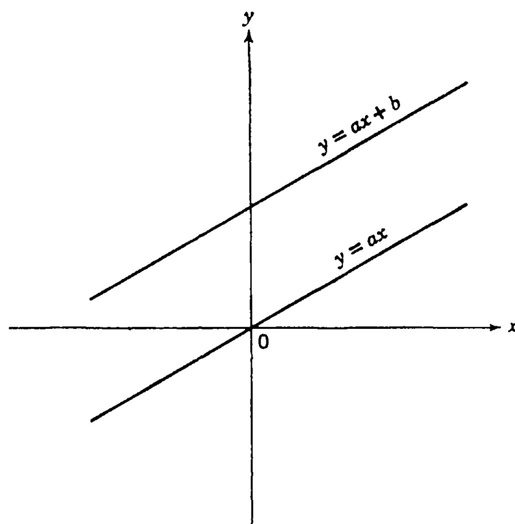


Figura 1.14 Funciones lineales.

La gráfica es una línea recta que pasa por el punto $x = 0$, $y = b$; la cual, si $a \neq 0$, pasa también por el punto $x = -b/a$, $y = 0$; y si $a = 0$ es una recta horizontal.

c) y es inversamente proporcional a x:

$$y = \frac{a}{x}.$$

En particular, para $a = 1$ es

$$y = \frac{1}{x},$$

de modo que

$$y = 1 \quad \text{para} \quad x = 1, \qquad y = 2 \quad \text{para} \quad x = \tfrac{1}{2}, \qquad y = \tfrac{1}{2} \quad \text{para} \quad x = 2.$$

La gráfica (ver Fig. 1.15) es una *hipérbola rectangular,* una curva simétrica con respecto a las bisectrices de los ángulos entre los ejes coordenados.

Esta función obviamente no está definida para el valor $x = 0$, puesto que la división por cero no tiene significado. En la vecindad del punto excepcional $x = 0$ la función tiene valores arbitrariamente grandes, tanto positivos como negativos; este es el ejemplo más simple de una *discontinuidad infinita,* un concepto que se discutirá posteriormente (ver p. 59).

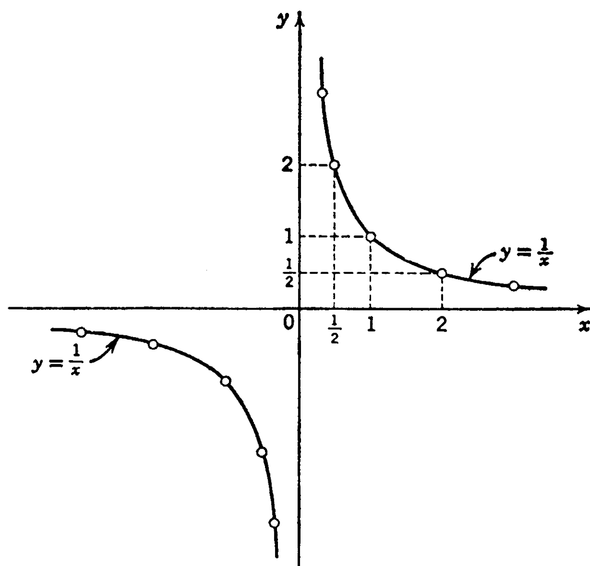


Figura 1.15 **Discontinuidad infinita.**

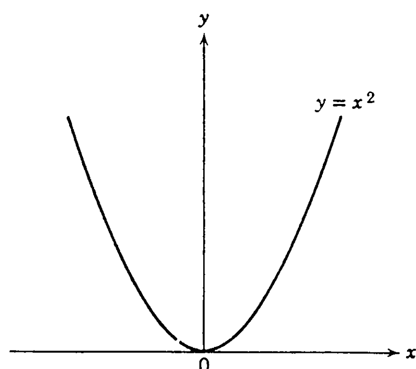


Figura 1.16 Parábola.

d) y es el cuadrado de x:

$$y = x^2.$$

Como es bien sabido, esta función es representada por una parábola (ver Fig. 1.16).

Análogamente, la función $y = x^3$ está representada por la así llamada *parábola cúbica* (ver Fig. 1.17).

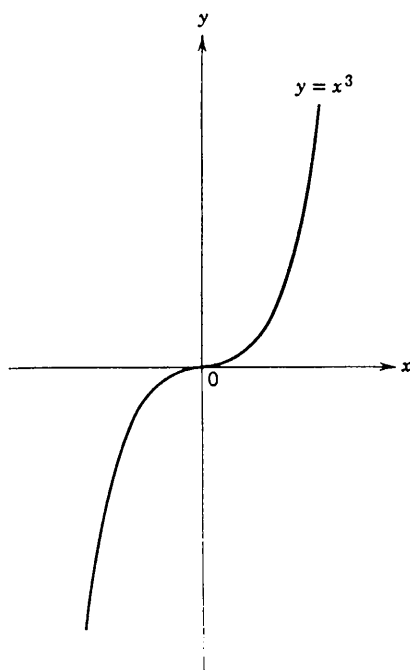


Figura 1 17 Parábola cúbica.

Funciones monótonas

Una función que para todos los valores de x en un intervalo tiene el mismo valor $y = a$ se denomina *función constante,* o simplemente *constante;* está representada gráficamente por una línea recta horizontal. Una función $y = f(x)$ para la cual un incremento en el valor de x resulta siempre en un incremento en el valor de y, esto es, para la cual $f(x) < f(x')$ siempre que $x < x'$, se denomina función *monótona creciente;* si, por otra parte, un incremento en el valor de x implica siempre un decremento en el valor de y, la función se denomina función *monótona decreciente.* Tales funciones son representadas gráficamente por curvas que siempre ascienden o bien siempre descienden conforme x recorre el intervalo de definición hacia valores crecientes (ver Fig. 1.18). Una función monótona transforma siempre valores distintos de x en diferentes y; esto es, la transformación es biunívoca o "uno a uno".

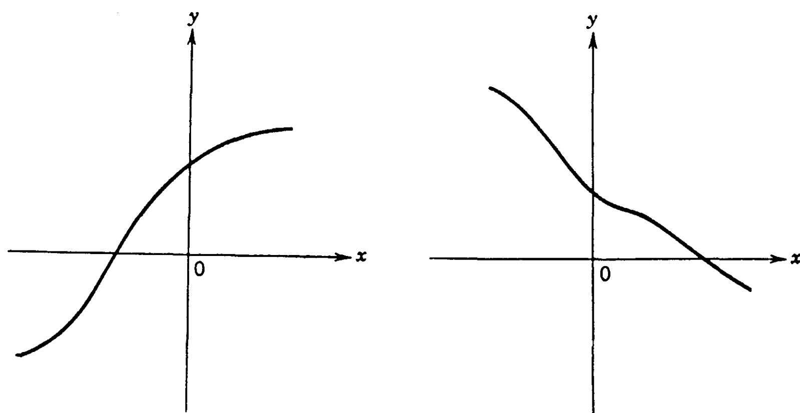


Figura 1.18 Funciones monótonas.

Funciones pares y funciones impares

Si la curva representada por $y = f(x)$ es simétrica con respecto al eje y, esto es, si $x = -a$ y $x = a$ dan el mismo valor de la función,

$$f(-x) = f(x),$$

la función se denomina función *par.* Por ejemplo, la función $y = x^2$ es par (ver Fig. 1.16). Si, por otra parte, la curva es simétrica con respecto al origen, esto es, si

$$f(-x) = -f(x),$$

se dice que la función es una función *impar;* entonces las funciones $y = x$, $y = x^3$ (ver Fig. 1.17), e $y = 1/x$ (ver Fig. 1.15) son impares.

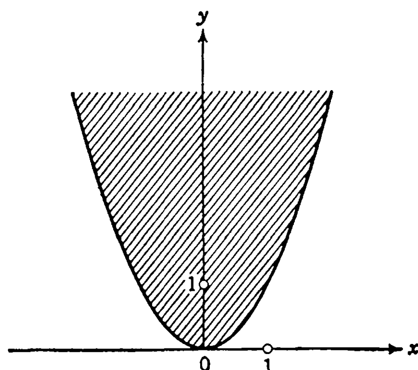


Figura 1.19 Gráfica de $y > x^2$.

Frecuentemente es útil considerar la representación geométrica de una desigualdad. Por ejemplo, la desigualdad $y > x^2$ está representada por el dominio sobre la parábola $y = x^2$ (Fig. 1.19). El interior del círculo unitario con centro en el origen (Fig. 1.20) se expresa por la desigualdad $x^2 + y^2 < 1$.

Frecuentemente, diversas desigualdades describen regiones más complicadas cuyas fronteras consisten de diferentes secciones. Así, el "primer" cuadrante del círculo unitario se describe por el sistema de desigualdades simultáneas:

$$x^2 + y^2 < 1, \qquad x > 0, \qquad y > 0.$$

(Ver Fig. 1.21.)

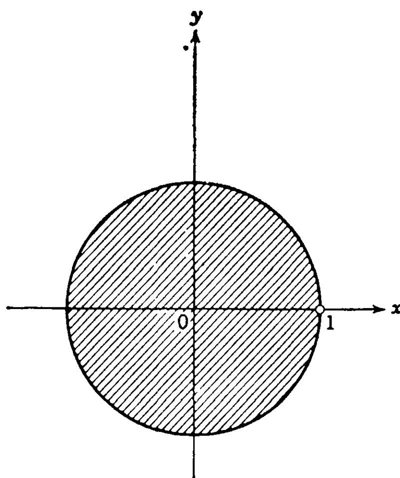


Figura 1.20 Gráfica de $x^2 + y^2 < 1$.

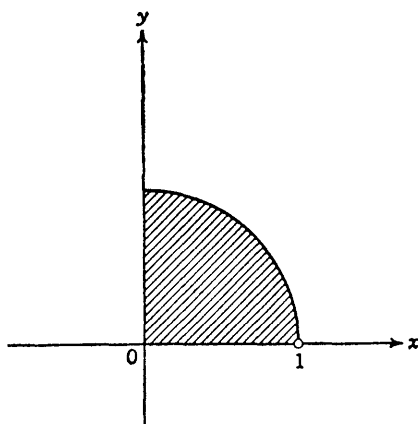


Figura 1.21 Gráfica de $x^2 + y^2 < 1$, $x > 0$, $y > 0$.

d. Continuidad

Explicaciones intuitiva y precisa

Las funciones y gráficas que se acaban de considerar exhiben una propiedad de gran importancia en el cálculo, a saber, la *continuidad*. Intuitivamente, la continuidad significa que un pequeño cambio en la variable independiente x implica sólo un pequeño cambio en la variable dependiente $y = f(x)$ y excluye saltos en el valor de y; así, la gráfica consiste de un solo pedazo de curva. En contraste, una gráfica $y = f(x)$ que consiste de pedazos de curva separados por un vacío en una abscisa x_0 exhibe allí una *discontinuidad de salto*. Por ejemplo, la función [1] $f(x) = \operatorname{sgn} x$, definida por $f(x) = +1$ para $x > 0$, por $f(x) = -1$ para $x < 0$ y $f(0) = 0$ tiene una "discontinuidad de salto" [2] en $x_0 = 0$ (ver Fig. 1.22).

La idea de la continuidad está implícita en el uso diario de las matemáticas elementales. Siempre que una función $y = f(x)$ esté descripta por tablas, tales como las logarítmicas o trigonométricas, el valor y puede ser registrado solamente para un conjunto "discreto" de valores de la variable independiente x; por ejemplo, a intervalos de $1/1000$ o bien $1/100{,}000$. Sin embargo, valores no registrados de la función pueden necesitarse para x intermedios. Entonces, tácitamente se supone que un valor no registrado $f(x_0)$ es aproximadamente el mismo que el de $f(x)$ para una x vecina que aparece en la tabla, y que $f(x_0)$ puede ser aproximado precisamente como

[1] Se pronuncia "signum" o bien "signo" de x.

[2] Técnicamente, la palabra "salto" se refiere solamente a la clase particular de discontinuidad en la cual la función aproxima desde la derecha e izquierda valores que no coinciden simultáneamente con $f(x_0)$. Una discontinuidad "infinita" es exhibida por la función $y = 1/x$ para $x \neq 0$ y $y = 0$ para $x = 0$. Aún otros tipos de discontinuidades se discutirán posteriormente.

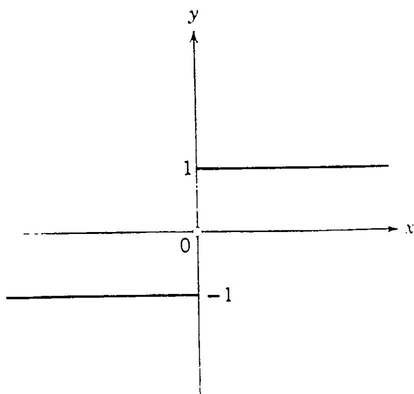


Figura 1.22 La función $f(x) = \operatorname{sgn} x$.

se quiere si sólo los valores de x en la tabla se espacian suficientemente cerca uno con otro.

La continuidad de la función $f(x)$ para un valor x_0 significa justamente que $f(x)$ difiere arbitrariamente poco del valor $f(x_0)$ cuando x está suficientemente cerca de x_0. Las expresiones "difiere arbitrariamente poco" y "suficientemente cerca" son un tanto vagas y deben ser explicadas precisamente en términos cuantitativos.

Prescríbase cualquier "margen de precisión" o "tolerancia", esto es, cualquier número real positivo ε (por pequeño que sea). Para la continuidad de f en x_0 se requiere que la diferencia entre $f(x)$ y $f(x_0)$ permanezca dentro de este margen, esto es, que $|f(x) - f(x_0)| < \varepsilon$, para todos los valores de x que se encuentran suficientemente cercanos a x_0 (o bien, para todos los valores x situados alguna distancia δ desde x_0).

Es más fácil visualizar lo que la continuidad significa si se interpreta a f como una transformación que asigna a puntos x en el eje x imágenes en

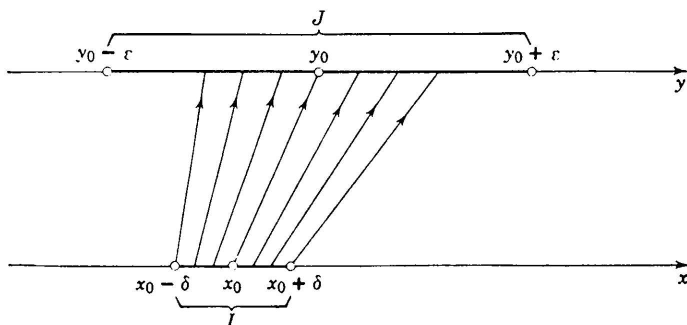


Figura 1.23 Continuidad del mapeo $y = f(x)$ en el punto x_0.

el eje y. Tómese cualquier punto x_0 en el eje x y su imagen $y_0 = f(x_0)$ (ver Fig. 1.23). Marcamos un intervalo abierto arbitrario J en el eje y, que tiene al punto y_0 como centro. Si la longitud de J es 2ε, entonces los puntos y de J son aquéllos cuya distancia a y_0 es menor que ε, o sea, para los cuales $|y - y_0| < \varepsilon$. La condición para la continuidad de $f(x)$ en x_0 es: Todos los puntos x suficientemente cercanos a x_0 tienen imágenes que se encuentran en J. O bien: Es posible marcar un intervalo I en el eje x con centro x_0, por ejemplo, el intervalo $x_0 - \delta < x < x_0 + \delta$, tal que todo punto x de I tiene una imagen $f(x)$ que se encuentra en J y por lo tanto $|f(x) - f(x_0)| < \varepsilon$. La continuidad de $f(x)$ en el punto x_0 significa que para una *ε-vecindad* arbitraria J del punto $y_0 = f(x_0)$ en el eje y, puede encontrarse una δ-vecindad I del punto x_0 en el eje x, cuya totalidad de puntos son transformados en puntos de J.[1] Por supuesto, esto solamente tiene sentido para puntos en el eje x en los cuales la función está definida, esto es, los que pertenecen al *dominio* de f. Lo anterior nos conduce a la siguiente *definición precisa de continuidad.*

La función $f(x)$ es continua en un punto x_0 de su dominio si para todo ε positivo puede encontrarse un número positivo δ tal que

$$|f(x) - f(x_0)| < \varepsilon$$

para todos los valores x en el dominio de f para los cuales $|x - x_0| < \delta$.

Muy útil es la interpretación geométrica de la continuidad cuando la función f se representa por su gráfica en el plano xy (ver Fig. 1.24). Sea $P_0 = (x_0, y_0)$ un punto sobre la gráfica. Los puntos (x, y) con $y_0 - \varepsilon < y < y_0 + \varepsilon$ forman ahora una "franja" horizontal J que contiene a P_0. La continuidad de f en x_0 significa que dada cualquiera de tales franjas horizontales J, no importa cuán angosta, es posible encontrar una franja vertical I dada por $x_0 - \delta < x < x_0 + \delta$ tan angosta que todo punto de la gráfica que se encuentre en I también caiga en J.

Como ilustración, considérese la función lineal $f(x) = 5x + 3$. Se tiene

$$|f(x) - f(x_0)| = |(5x + 3) - (5x_0 + 3)| = 5\,|x - x_0|,$$

la cual expresa que la transformación $y = 5x + 3$ *magnifica* distancias por el factor 5. Obviamente aquí $|f(x) - f(x_0)| < \varepsilon$ para todos los x para los

[1] En esta definición de continuidad I y J son intervalos que tienen sus *centros* en los puntos x_0 e y_0, respectivamente. Esto es conveniente para la definición analítica de continuidad en x_0 la cual recurre a las distancias $|x - x_0|$ e $|y - y_0|$; pero es algo artificial si se interpreta f geométricamente como un "mapeo" o transformación. En su lugar, se puede definir igualmente bien la continuidad de $y = f(x)$ en un punto x_0 por el requisito de que para todo intervalo *abierto* J en el eje y que contiene al punto $y_0 = f(x_0)$ puede encontrarse un intervalo abierto I en el eje x conteniendo al punto x_0 y tal que la imagen y de cualquier punto x en I para el cual la función esté definida está situada en J. La demostración de la equivalencia de las dos definiciones se deja al lector como un sencillo ejercicio.

cuales $|x - x_0| < \varepsilon/5$. Consecuentemente, la condición para la continuidad de $f(x)$ en el punto x_0 se satisface si se elige $\delta = \varepsilon/5$ (pero, por supuesto, cualquier número positivo $\delta < \varepsilon/5$ es también una posible elección); y la imagen de cualquier punto del intervalo $x_0 - \delta < x < x_0 + \delta$ se encontrará entonces en el intervalo $y_0 - \varepsilon < y < y_0 + \varepsilon$. En este ejemplo, a la aseveración de que la distancia $|y - y_0|$ es "arbitrariamente pequeña" para $|x - x_0|$ "suficientemente pequeño" puede dársele un significado específico en efecto, $|x - x_0|$ es *suficientemente pequeño* si no excede un quinto del valor de $|y - y_0|$.

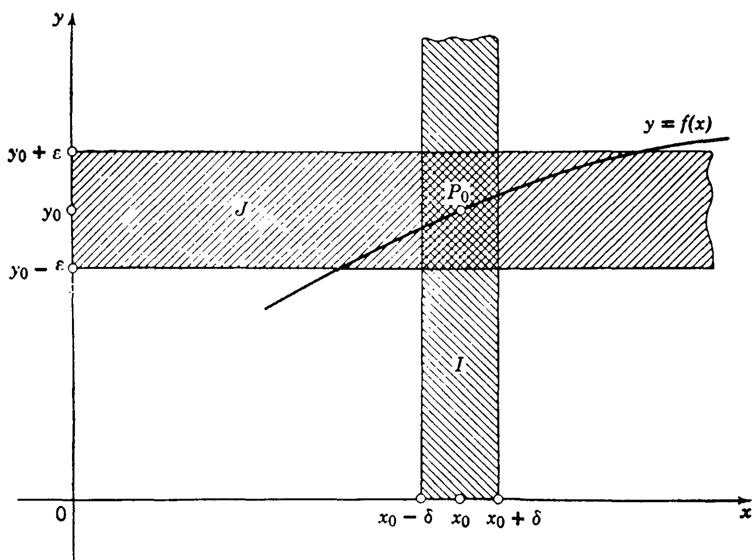


Figura 1.24 Continuidad de $y = f(x)$ en el punto x_0.

Otro ejemplo es proporcionado por la función $f(x) = x^2$. Aquí se tiene para $|x - x_0| < \delta$ que

$$\begin{aligned} |f(x) - f(x_0)| &= |x^2 - x_0^2| = |x - x_0|\,|2x_0 + (x - x_0)| \\ &\leq |x - x_0|(2\,|x_0| + |x - x_0|) \leq \delta(2\,|x_0| + \delta). \end{aligned}$$

Se verifica inmediatamente que la condición $|f(x) - f(x_0)| < \varepsilon$ se satisface si se elige $\delta = -|x_0| + \sqrt{\varepsilon + |x_0|^2}$.

Intuitivamente, la idea de la continuidad parece obvia sin explicación, pero la formulación precisa puede ser en un principio algo difícil de entender debido a la vaguedad de palabras tales como "uno puede encontrar" o bien "elegido arbitrariamente". Sin embargo, el lector, que puede estar al principio muy satisfecho con alguna noción intuitiva de la continuidad, aprenderá gradualmente a apreciar la lógica precisión y generalidad de la

definición analítica, resultado de un largo y persistente esfuerzo para conciliar la necesidad del entendimiento intuitivo con el de la claridad lógica. Es indispensable un significado preciso para la palabra "continuidad". La definición analítica dada aquí es la formulación requerida y precisa de una importante propiedad de las funciones.

Para el principiante se deberá nuevamente insistir en que "pequeño" no es una designación absoluta de un número; más bien, el término "arbitrariamente pequeño" se refiere a un número que no es fijo al principio, sino que, cualquier valor positivo ε puede ser seleccionado, el cual está sujeto a una elección subsecuente menor para una aproximación refinada de $f(x_0)$. "Suficientemente pequeño" se refiere a un número δ que debe ser ajustado para satisfacer un margen de tolerancia establecido anteriormente por otro número ε.

Explicación de la continuidad y de la discontinuidad mediante ejemplos. Podemos aclarar la definición de continuidad haciendo contraste con ejemplos de discontinuidad, ejemplos que no satisfacen la definición anterior. Recuérdese el ejemplo sencillo de la función $f(x) = \operatorname{sgn} x$ en la p. 55. Obviamente, para todo $x_0 \neq 0$ esta función es continua de acuerdo con la definición en términos de ε y δ; de hecho, con un $\delta = |x_0|$ constante, no importa cuán pequeño se elija ε. Pero para $x_0 = 0$ no puede encontrarse ninguna δ si ε es menor que 1, puesto que $|f(x) - f(0)| = |f(x)| = 1 > \varepsilon$ para todo x distinto de cero, sin importar cuán próximo está x de cero.

La función $\operatorname{sgn} x$ ilustra el tipo más simple de discontinuidad en un punto ξ, conocida como *discontinuidad de salto,* en la cual $f(x)$ se aproxima a valores límite desde la derecha e izquierda, conforme x se aproxima a valores límite ξ, que difieren, ya sea el uno del otro o bien del valor de f en el punto ξ.[1] La gráfica en $x = \xi$ tiene entonces un vacío. Otras curvas con discontinuidades de salto están bosquejadas en la Fig. 1.25 *a* y *b*; la definición de estas funciones deberá resultar evidente de las figuras.[2]

En discontinuidades de este género existen tanto los límites por la derecha como por la izquierda. Volvamos ahora a discontinuidades en las cuales éste no es el caso. Las más importantes de éstas son las *discontinuidades infinitas* o infinidades. Estas son discontinuidades análogas a las exhibidas

---

[1] La definición precisa de *límite* se dará en la Sección 1.7; para los comentarios descriptivos que se hacen aquí, una idea intuitiva es suficiente.

[2] En todos estos ejemplos de discontinuidades de salto, los límites de la función por la derecha y por la izquierda en el punto de discontinuidad poseen valores diferentes. El ejemplo trivial de la función $f(x)$ definida por

$$f(x) = 0 \quad \text{para} \quad x \neq 0, \qquad f(x) = 1 \quad \text{para} \quad x = 0$$

ilustra una discontinuidad de salto en la cual los límites desde ambos lados son iguales entre sí pero difieren del valor de f en el punto mismo ξ de discontinuidad. Se tiene entonces una singularidad *eliminable.* Aquí f puede hacerse continua cambiando solamente el valor de f en ξ de modo que coincida con los límites desde ambos lados.

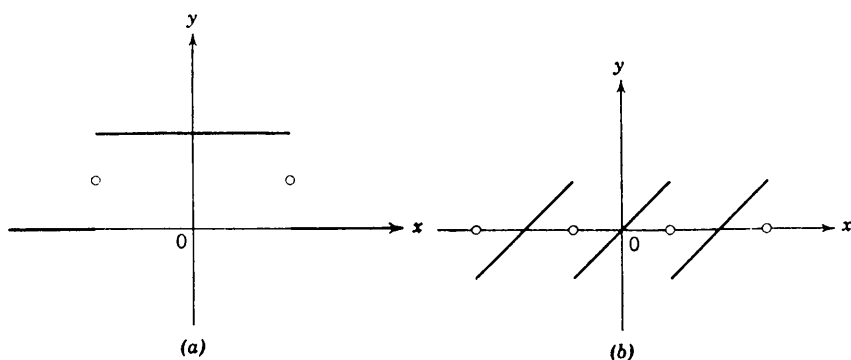


Figura 1.25.

por las funciones $1/x$ y $1/x^2$ en el punto $x = 0$; conforme $x \to 0$, el valor absoluto $|f(x)|$ de la función crece más allá de toda cota. La función $1/x$ crece numéricamente más allá de toda cota a través de valores positivos y negativos respectivamente, conforme x se aproxima al origen desde la izquierda y desde la derecha. Por otra parte, la función $1/x^2$ tiene para $x = 0$ una discontinuidad infinita en la cual el valor de la función crece

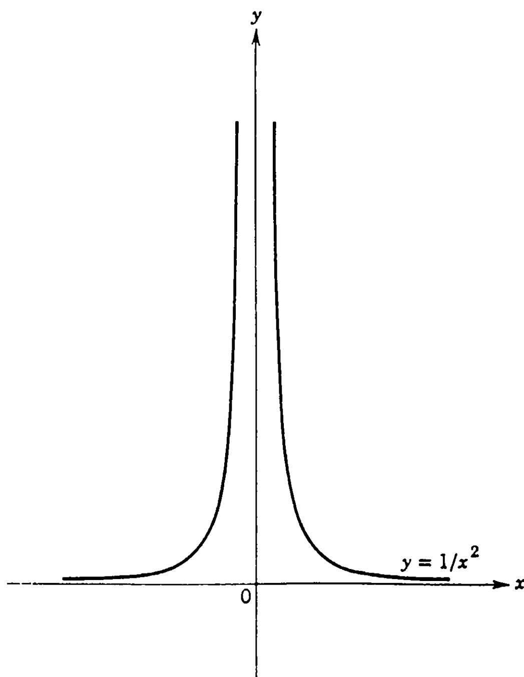


Figura 1.26 Gráfica de función con discontinuidad infinita.

más allá de cualquier cota positiva conforme x se aproxima al origen desde ambos lados (ver Fig. 1.26 y Fig. 1.27). La función $1/(x^2 - 1)$ mostrada en la Fig. 1.27 posee discontinuidades infinitas tanto en $x = 1$ cocmo en $x = -1$.

Un ejemplo de otro tipo de discontinuidad en el cual no existe ningún límite desde la derecha o desde la izquierda es la función par "lineal por pedazos", $y = f(x)$, ilustrada en la Fig. 1.28, la cual está definida para todos los valores de x distintos de cero como sigue. Esta función toma alternativamente los valores $+1$ y -1 para los valores $-x$ de la forma $\pm 1/2^n$,

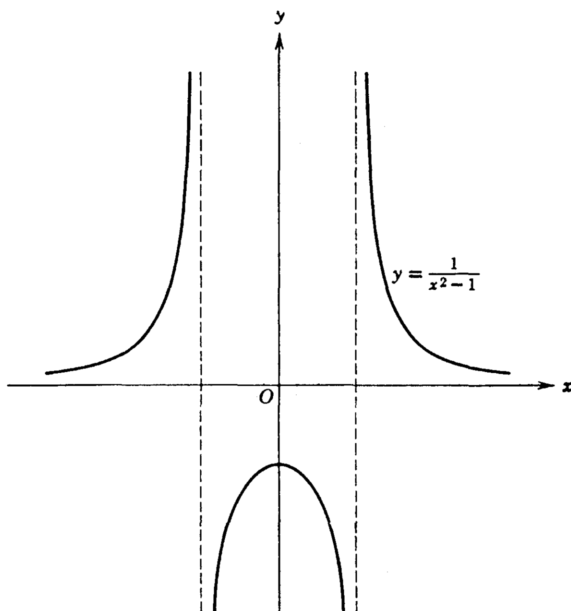


Figura 1.27 Función con discontinuidades infinitas.

donde n es cualquier entero: $f(\pm 1/2^n) = (-1)^n$. En todo intervalo $1/2^{n+1} < x < 1/2^n$, o bien $-1/2^n < x < -1/2^{n+1}$, la función $f(x)$ es lineal y recorre todos los valores entre -1 y $+1$. Por ello la función oscila hacia arriba y hacia abajo cada vez con más rapidez entre los valores -1 y $+1$ conforme x se aproxima más y más al punto $x = 0$, y en la vecindad inmediata de ese punto ocurren un número infinito de tales oscilaciones. Un comportamiento análogo es exhibido por la curva suave de la Fig. 1.29. [Aquí, de hecho $f(x)$ está dada por una expresión en forma exacta, a saber, $f(x) = \text{sen}\,(1/x)$, con la función seno definida apropiadamente como en la p. 74.]

Un contraste con este ejemplo es la función lineal por pedazos $y = f(x)$ que toma los valores $f(\pm 1/2^n) = (-\frac{1}{2})^n$ para todos los enteros n (ver

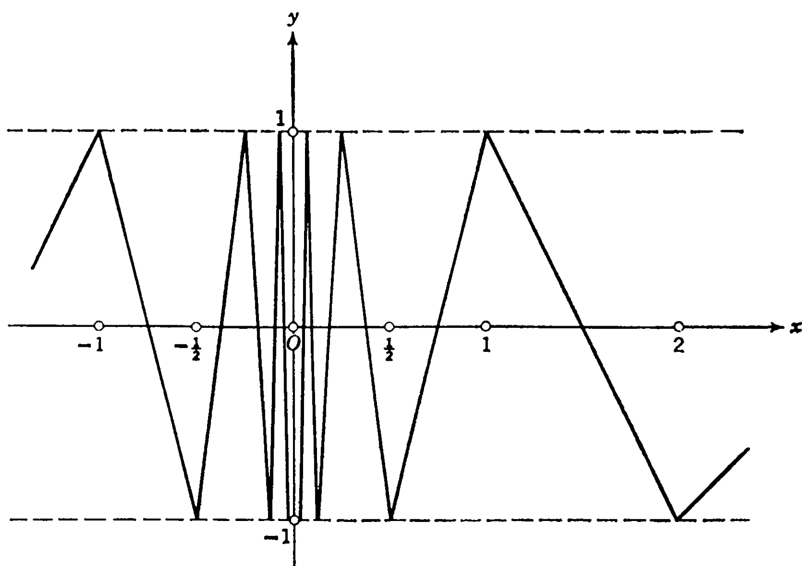


Figura 1.28 Función oscilante, lineal por pedazos, con discontinuidad.

Fig. 1.30) y es lineal para valores intermedios x. Aquí $f(x)$ permanece continua en el punto $x = 0$ si le asignamos el valor 0 en este punto. En la vecindad del origen la función oscila hacia arriba y hacia abajo un número infinito de veces, pero la magnitud de estas oscilaciones se hace arbitra-

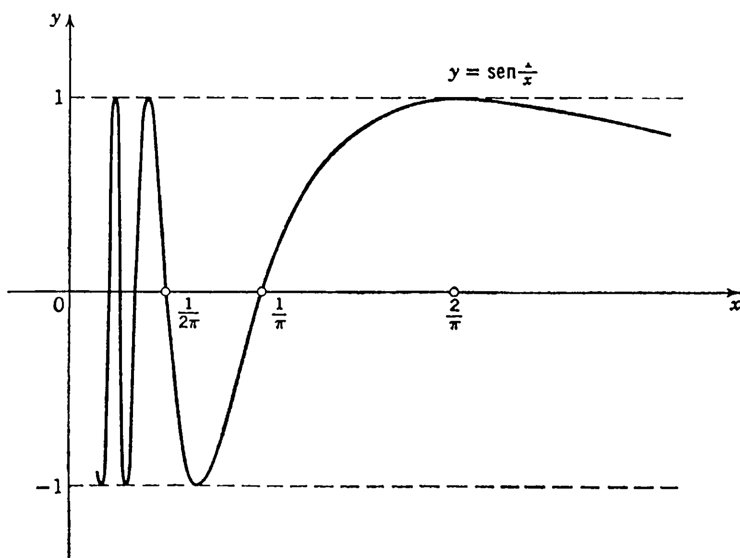


Figura 1.29 Función oscilante con discontinuidad.

riamente pequeña conforme se aproximan al origen. La situación es la misma para la función $y = x \operatorname{sen}(1/x)$ (ver Fig. 1.31).

Estos ejemplos muestran que la continuidad permite toda clase de posibilidades notables ajenas a nuestra cándida intuición.

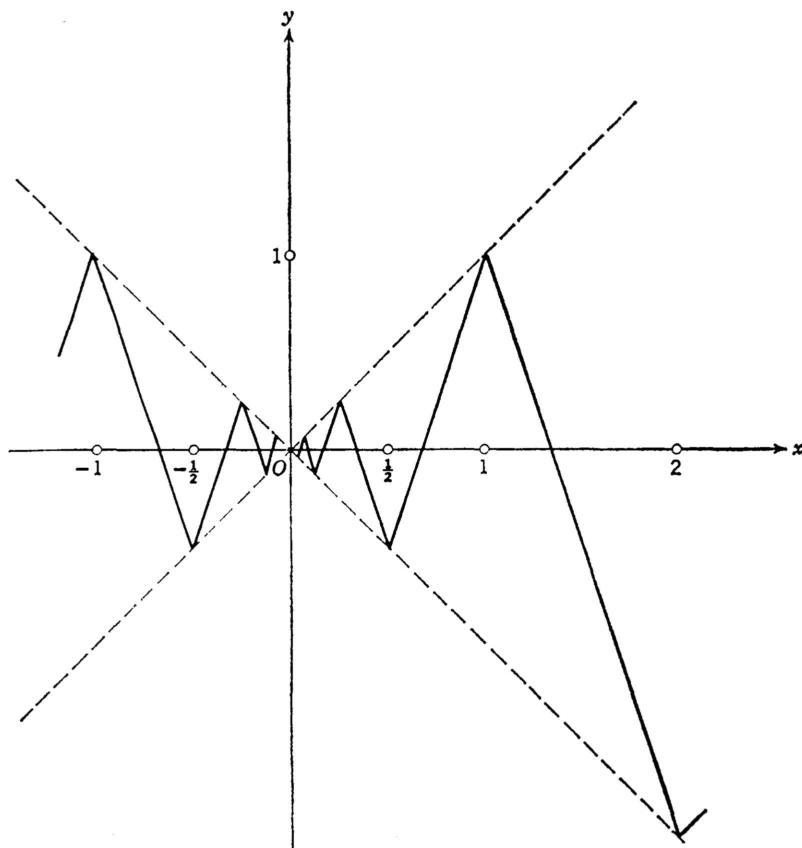


Figura 1.30 Función continua oscilante.

**Discontinuidades removibles*

Como se habrá notado, puede suceder que en un cierto punto, digamos $x = 0$, una función no esté definida por la ley original, tal como en los últimos ejemplos discutidos. Estamos entonces en libertad de extender la definición de la función, asignándole cualquier valor deseado en tal punto. En el último ejemplo podemos escoger la definición de tal manera que la función se torne también *continua* en ese punto, a saber, eligiendo $y = 0$ en $x = 0$. Una extensión continua análoga puede ser definida siempre que el límite por la izquierda así como el límite por la derecha existan

y sean iguales; bastará entonces con hacer el valor de la función en el punto en cuestión igual a estos límites de manera de hacer continua la función allí. Cualquiera que sea la discontinuidad que pueda ser impuesta por definición en $x = 0$, esta discontinuidad es "eliminable" asignando un valor adecuado a $f(0)$. Para la función $y = \text{sen } 1/x$, como para la función de la Fig. 1.28, esto, sin embargo, no es posible: cualquiera que sea el

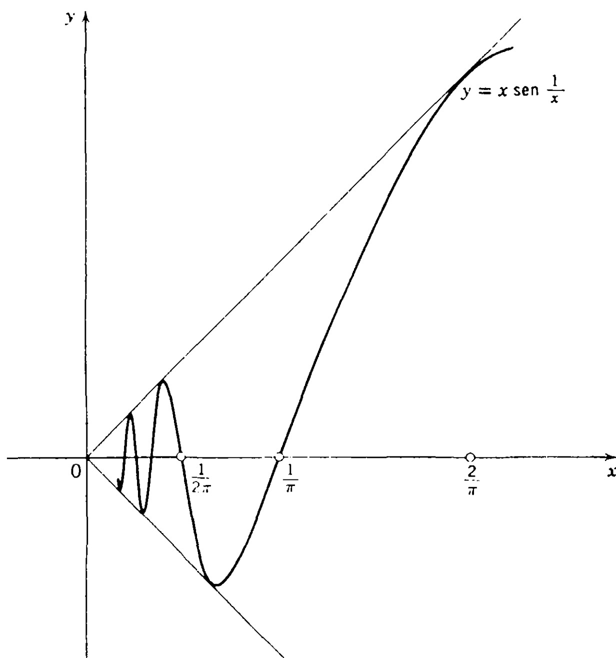


Figura 1.31 Función continua oscilante.

valor que se asigne a la función en $x = 0$, la función extendida es discontinua.

Módulo de continuidad. Continuidad uniforme. Nuestra definición de continuidad de la función $f(x)$ en x_0 requiere que para toda medida de precisión, $\varepsilon > 0$, *existan* cantidades $\delta > 0$ (llamadas módulos de continuidad) tales que $|f(x) - f(x_0)| < \varepsilon$ para todos los valores de x en el dominio de f para los cuales $|x - x_0| < \delta$. Un módulo de continuidad proporciona información sobre la sensibilidad de f a los cambios en x. Un módulo de continuidad δ nunca es único; puede ser siempre reemplazado (para las mismas x_0 y ε) por cualquier cantidad positiva menor δ', puesto que $|x - x_0| < \delta'$ implica $|x - x_0| < \delta$ y por ello $|f(x) - f(x_0)| < \varepsilon$. Para fines prácticos, como en los cálculos numéricos, podemos estar interesados

en una elección particular de δ; por ejemplo, en el máximo valor para δ. Por otra parte, si meramente deseamos establecer el hecho de que f es continua en x_0, entonces solamente se requiere exhibir un módulo cualquiera de continuidad para todo ε positivo.

En general, como lo muestran los ejemplos, este $\delta = \delta(\varepsilon)$ depende no sólo de ε sino que también del valor de x_0. Por supuesto, no se requiere considerar todos los valores positivos posibles de ε. Pueden restringirse siempre las consideraciones a ε suficientemente pequeños; digamos, a $\varepsilon \leq \varepsilon_0$ para un ε_0 elegido arbitrariamente, puesto que para $\varepsilon > \varepsilon_0$ puede utilizarse el mismo módulo de continuidad que para $\varepsilon = \varepsilon_0$. Análogamente, debemos tomar en cuenta solamente los puntos x en el dominio de f situados en una vecindad arbitraria de x_0; digamos, aquellos puntos con $|x - x_0| < \delta_0$, puesto que puede siempre reemplazarse cualquier módulo de continuidad δ por uno menor que no exceda a δ_0. La continuidad de f en x_0 es una *propiedad local,* es decir, una propiedad que depende solamente de los valores de f en alguna vecindad de x_0 arbitrariamente pequeña.

Como se ha visto, la función f puede ser continua para algún x_0 y discontinua para otros x. Una función se denomina *continua en un intervalo* si es continua en cada punto del intervalo. Para cada x_0 del intervalo se tiene entonces un módulo de continuidad $\delta = \delta(\varepsilon)$ el cual puede esperarse que varíe con x_0, reflejando las diferentes proporciones según las cuales y varía al variar x cerca de puntos x_0 diferentes.

Se dice que f es *uniformemente continua* en un intervalo si puede encontrarse un *módulo uniforme de continuidad* $\delta = \delta(\varepsilon)$ para ese intervalo, esto es, uno que no dependa del punto particular x_0 del intervalo. Por lo tanto, $f(x)$ es uniformemente continua en un intervalo[1] si para cada ε positivo existe un número positivo δ tal que $|f(x) - f(x_0)| < \varepsilon$ para *dos* puntos *cualesquiera* x y x_0 del intervalo para los cuales $|x - x_0| < \delta$.

Para una función uniformemente continua $y = f(x)$ los valores de y difieren "arbitrariamente poco" uno del otro para cualesquiera valores x que estén "suficientemente cercanos", *independientemente de su posición en el intervalo.* En ciertos aspectos la continuidad uniforme se acerca más a nociones intuitivas que la mera propiedad local de continuidad.

Por ejemplo, la función $f(x) = 5x + 3$ es uniformemente continua para todos los valores de la variable independiente puesto que aquí $|f(x) - f(x_0)| = 5|x - x_0| < \varepsilon$ para $|x - x_0| < \varepsilon/5$, y, por lo tanto, $\delta(\varepsilon) = \varepsilon/5$ representa un módulo uniforme de continuidad.

La función $f(x) = x^2$ para un intervalo x infinito, ciertamente no es uniformemente continua. Es evidente que pequeñas variaciones en x pueden producir cambios arbitrariamente grandes en x^2, bastando con que x sea suficientemente grande. Una ojeada a una tabla de cuadrados de enteros

[1] En esta definición la palabra *intervalo* puede referirse ya sea a intervalos cerrados o abiertos o bien infinitos.

x muestra cómo los cuadrados sucesivos se encuentran espaciados más y más conforme x crece. Pero, sin embargo, si se consideran solamente pares de valores x y x_0 pertenecientes a un intervalo fijo finito cerrado $[a, b]$, puede encontrarse un módulo uniforme de continuidad. En efecto, para $|x - x_0| < \delta$ se tiene

$$|f(x) - f(x_0)| = |x^2 - x_0^2| = |x - x_0|\,|x + x_0| \leq 2\,|x - x_0|(|b| + |a|)$$
$$< 2\delta(|b| + |a|) = \varepsilon$$

si se toma $\delta = \varepsilon/2(|b| + |a|)$.

Prevalece una situación análoga para la función $f(x) = 1/x$ para $x \neq 0$, con $f(0) = 0$. Considérese un intervalo cerrado acotado $a \leq x \leq b$ en todo el cual la función es continua. Tal intervalo no puede incluir al origen, el cual es un punto de discontinuidad, de modo que a y b deberán tener el mismo signo. Supóngase que a y b son ambos positivos. Entonces, para x y x_0 pertenecientes al intervalo y para $|x - x_0| < \delta$ se tiene

$$\left|\frac{1}{x} - \frac{1}{x_0}\right| = |x_0 - x|\,\frac{1}{|x_0|\,|x|} < \frac{\delta}{a^2} = \varepsilon$$

para $\delta = a^2\varepsilon$. Por lo tanto, la función es uniformemente continua en el intervalo $[a, b]$. Por supuesto, esto prueba también que la función $f(x) = 1/x$ es continua en todo punto $x_0 > 0$. Puesto que todo tal valor x_0 puede ser incluido en algún intervalo $a < x_0 < b$ con a y b positivos. La expresión $\delta = a^2\varepsilon$ es entonces un módulo de continuidad para la función en x_0, si se restringe x a una vecindad de x_0 que se encuentre totalmente en el intervalo.

Salta a la vista que las funciones continuas en los ejemplos anteriores son uniformemente continuas en cualquier intervalo cerrado acotado que pertenece a su dominio. Estas ilustran una propiedad general que será demostrada en el Suplemento, p. 123. Podemos enunciarla:

Toda función continua en un intervalo cerrado y acotado es por necesidad uniformemente continua en ese intervalo.

La restricción a intervalos *acotados* es esencial como lo muestra el ejemplo de la función x^2. Análogamente, se deberá estipular que el intervalo es *cerrado*. Por ejemplo, la función $y = 1/x$ es continua en el intervalo abierto $0 < x < 1$, pero no es uniformemente continua allí; pues cambios arbitrariamente grandes en y pueden ser producidos por cambios arbitrariamente pequeños en x, bastando sólo con que x se encuentre suficientemente cercano al origen. Si existiera un módulo uniforme de continuidad $\delta(\varepsilon)$ para el intervalo $(0, 1)$, podría tomarse, por ejemplo, $x_0 < \delta$, $x = \frac{1}{2}x_0$; obviamente entonces $|1/x - 1/x_0| = 1/x_0$ es mayor que cualquier ε preasignado siempre que x_0 sea suficientemente pequeño, de modo que la suposición de un $\delta(\varepsilon)$ uniforme conduce a una contradicción.

Continuidad de Lipschitz—Continuidad de Hölder. En los ejemplos anteriores de funciones uniformemente continuas en un intervalo $[a, b]$, se encontró un módulo de continuidad particularmente simple, a saber, $\delta(\varepsilon)$ proporcional a ε. Esta situación muy común se presenta en las así llamadas funciones *"continuas de Lipschitz"*,* esto es, las funciones $f(x)$ que satisfacen una desigualdad de la forma

$$|f(x_2) - f(x_1)| \leq L\,|x_2 - x_1|$$

(la llamada *condición de Lipschitz)* para todos los x_1 y x_2 en el intervalo, con un valor fijo L. La continuidad de Lipschitz significa que el "cociente incremental"

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

formado para dos puntos distintos cualesquiera del intervalo nunca excede un valor finito L, fijo, en valor absoluto; o bien que la transformación $y = f(x)$ magnifica distancias de puntos en el eje x a lo más por el factor L. Evidentemente, para una función "continua de Lipschitz" la expresión $\delta(\varepsilon) = \varepsilon/L$ es un módulo de continuidad puesto que $|f(x_2) - f(x_1)| < \varepsilon$ para $|x_2 - x_1| < \varepsilon/L$. Recíprocamente, cualquier función con un módulo de continuidad proporcional a ε, digamos, $\delta(\varepsilon) = c\varepsilon$, es "continua de Lipschitz", con $L = 1/c$.

Como se verá en el capítulo 2, la mayoría de las funciones tratadas serán "continuas de Lipschitz" excepto en puntos aislados, como una consecuencia del hecho que sus derivadas son acotadas en cualquier intervalo cerrado que excluya a estos puntos. Sin embargo, la continuidad de Lipschitz es solamente suficiente pero no necesaria para la continuidad uniforme. El ejemplo más simple de una función que es continua sin ser "continua de Lipschitz" está dada por $f(x) = \sqrt{x}$ para $x \geq 0$, y hay que considerar $x_0 = 0$. Aquí el cociente incremental

$$\frac{f(x) - f(0)}{x - 0} = \frac{1}{\sqrt{x}}$$

se torna arbitrariamente grande para x suficientemente pequeños, y, por lo tanto, no puede ser acotado por una constante fija L. De aquí que no sea posible elegir $\delta(\varepsilon)$ proporcional a ε; pero existen otros módulos de continuidad no lineales para esta función, por ejemplo $\delta(\varepsilon) = \varepsilon^2$.

La función $\sqrt{x}$ pertenece a la clase general de funciones llamadas "continuas de Hölder",** que satisfacen una "condición de Hölder":

$$|f(x_2) - f(x_1)| \leq L\,|x_2 - x_1|^\alpha$$

para todos x_1, x_2 de un intervalo, donde L y α son constantes fijas, estando el "exponente de Hölder" α restringido a valores $0 < \alpha \leq 1$. Las funciones continuas Lipschitz se obtienen para el exponente Hölder especial $\alpha = 1$.

---

* (N. del T.) Decimos así para abreviar la expresión *"continuas en el sentido de Lipschitz"*. También se llaman "Lipschitzianas".

** (N. del T.) Brevemente, en lugar de "continuas en el sentido de Höolder". También llamadas "Hölderianas".

Es claro que $\delta = L^{-1/\alpha}\varepsilon^{1/\alpha}$ es un posible módulo de continuidad para una función "continua de Hölder" f; aquí δ es proporcional a $\varepsilon^{1/\alpha}$, y no a ε. La función $f(x) = \sqrt{x}$ es "continua de Hölder" con exponente $\alpha = \frac{1}{2}$. Esto se sigue de la desigualdad

$$|\sqrt{x_2} - \sqrt{x_1}| \leq |x_2 - x_1|^{1/2},$$

la que se obtiene observando que

$$|\sqrt{x_2} - \sqrt{x_1}| \leq |\sqrt{x_2} + \sqrt{x_1}|$$

y multiplicando por $|\sqrt{x_2} - \sqrt{x_1}|$. Esto proporciona el módulo de continuidad $\delta(\varepsilon) = \varepsilon^2$ para $\sqrt{x}$, como se mencionó anteriormente.

Con más generalidad, las potencias fraccionales $f(x) = x^\alpha$ para $0 < \alpha \leq 1$ son "continuas de Hölder" con exponente de Hölder α.

Las funciones hölderianas aún no agotan la clase de todas las funciones uniformemente continuas. No es difícil construir ejemplos de funciones continuas para las cuales las potencias de ε no sirven como módulos de continuidad. (Ver problema 13, p. 140.)

e. El teorema del valor intermedio. Funciones inversas

Intuitivamente, no existe duda de que una función que es continua, y, de aquí, que no tiene "saltos", no puede variar de un valor a otro sin pasar por todos los valores intermedios. Este hecho se expresa por el así llamado teorema del valor intermedio (su demostración precisa se da en el Suplemento, p. 123). Lo enunciamos así:

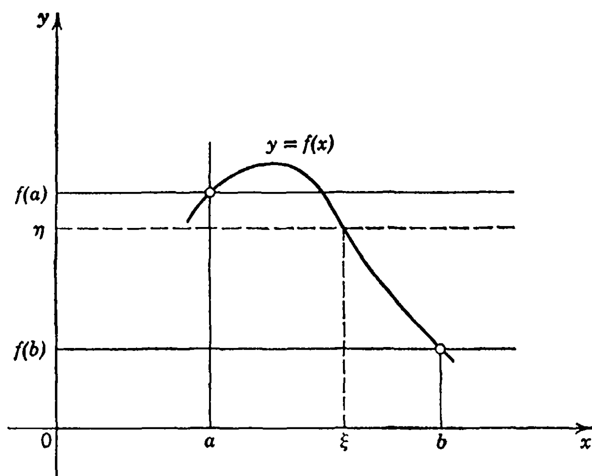


Figura 1.32 El teorema del valor intermedio.

TEOREMA DEL VALOR INTERMEDIO. *Considérese una función $f(x)$ continua en todo punto de un intervalo. Sean a y b dos puntos cualesquiera del intervalo y sea η cualquier número entre $f(a)$ y $f(b)$. Entonces, existe un valor ξ entre a y b para el cual $f(\xi) = \eta$.*

Interpretado geométricamente, el teorema establece que si dos puntos $(a, f(a))$ y $(b, f(b))$ de la gráfica de una función continua f están situados en diferentes lados de una paralela $y = \eta$ al eje x, entonces la paralela *intersecta* a la gráfica en algún punto intermedio (ver Fig. 1.32). Por supuesto que pueden haber varias intersecciones. En el importante caso en que la función $f(x)$ es monótona creciente o bien monótona decreciente en todo el intervalo, puede haber solamente una intersección, pues f no puede tener el mismo valor η para dos valores diferentes de ξ.

Como ejemplo tómese la función $f(x) = x^2$, la cual es monótona creciente y continua en el intervalo $1 \leq x \leq 2$. Aquí $f(1) = 1$ y $f(2) = 4$. Tomando para η el valor 2, intermedio entre 1 y 4, se encuentra que existe un único ξ entre 1 y 2 para el cual $\xi^2 = 2$. Este es, por supuesto, el número que se denota por $\sqrt{2}$.

Continuidad de la función inversa

Para cualquier función monótona creciente $f(x)$, definida en un intervalo $a \leq x \leq b$, se encuentra que para todo número real η tal que $f(a) \leq \eta \leq f(b)$, existe exactamente un ξ tal que $a \leq \xi \leq b$ para el cual $f(\xi) = \eta$.[1] Sean $\alpha = f(a)$ y $\beta = f(b)$. Puesto que ξ está determinado de manera única por η, se tiene una *función* $\xi = g(\eta)$ definida para argumentos η en el intervalo cerrado $[\alpha, \beta]$. Esta función g se denomina la *inversa* de f. Puesto que a valores más grandes de ξ corresponden valores más grandes de $\eta = f(\xi)$, la función g es también monótona creciente. Es fácil mostrar que la función inversa g también es continua.

En efecto, sea η cualquier valor entre α y β (ver Fig. 1.33). Entonces $\xi = g(\eta)$ debe estar situado entre $a = g(\alpha)$ y $b = g(\beta)$. Sea ε un número positivo dado, el cual puede suponerse tan pequeño que sea $a < \xi - \varepsilon < \xi + \varepsilon < b$. Se debe mostrar que $|g(y) - g(\eta)| < \varepsilon$ para toda y suficientemente próxima a η. Puesto que f es creciente, $\eta = f(\xi)$ está situado entre los valores $f(\xi - \varepsilon) = A$ y $f(\xi + \varepsilon) = B$, y puede encontrarse un δ tan pequeño que sea

$$A < \eta - \delta < \eta + \delta < B.$$

Si y es cualquier valor tal que $\eta - \delta < y < \eta + \delta$, y es $x = g(y)$, se tiene que $A < y < B$ y, por lo tanto, $g(A) < g(y) < g(B)$; esto es, $\xi - \varepsilon < g(y) < \xi + \varepsilon$, o bien $|g(y) - g(\eta)| < \varepsilon$. La misma demostración, ligeramente modificada, se aplica cuando η es uno de los puntos extremos α o β del intervalo de definición de g.

Las relaciones $y = f(x)$ y $x = g(y)$ son equivalentes y están representadas por la misma gráfica en el plano $\{x, y\}$; los puntos (x, y) del plano

---

[1] El teorema del valor intermedio, como se lo estableció, asigna ξ para η en el intervalo *abierto* $f(x) < \eta < f(b)$. Sin embargo, por supuesto que para $\eta = f(a)$, o bien para $\eta = f(b)$, basta sólo con tomar $\xi = a$, o bien $\xi = b$.

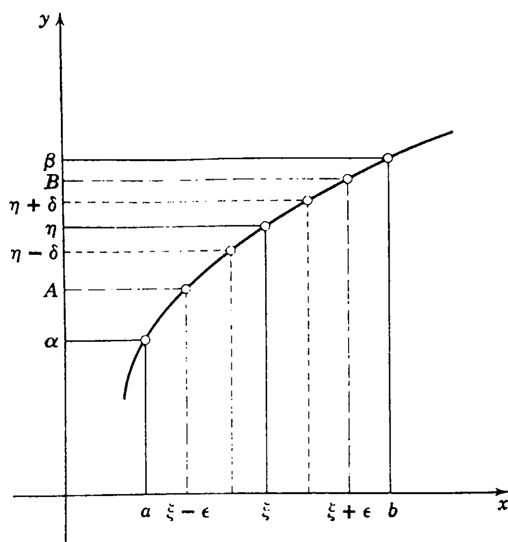


Figura 1.33 Continuidad de la inversa de una función monótona continua.

para los cuales $y = f(x)$ son los mismos que los puntos para los cuales $x = g(y)$. Si la función g se representa de la manera usual por $y = g(x)$, deben intercambiarse x e y; entonces la gráfica de $y = g(x)$ se obtiene de la gráfica $y = f(x)$ tomando la imagen especular respecto a la línea $y = x$.

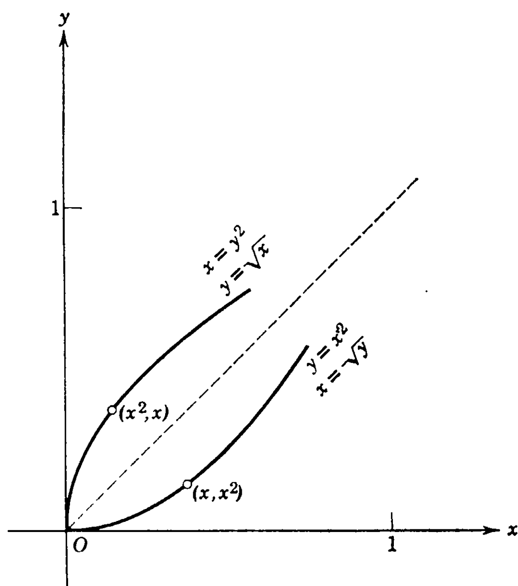


Figura 1.34 Funciones inversas.

Se tiene un ejemplo simple con las gráficas de la función $f(x) = x^2$ para $x \geq 0$ y de la función inversa $g(x) = \sqrt{x}$ para $x \geq 0$ (ver Fig. 1.34).

1.3 Las funciones elementales

a. Funciones racionales

Pasemos ahora a un breve repaso de las conocidas funciones elementales. Los tipos más simples de función se construyen mediante aplicación repetida de las operaciones elementales, adición y multiplicación. Si se aplican estas operaciones a una variable independiente x y a un conjunto de números reales $a_1, \ldots, a_n$ se obtienen los *polinomios*

$$y = a_0 + a_1x + \cdots + a_nx^n.$$

Los polinomios son las funciones más simples del análisis y, en cierto sentido, las básicas.

Cocientes de tales polinomios, de la forma

$$y = \frac{a_0 + a_1x + \cdots + a_nx^n}{b_0 + b_1x + \cdots + b_mx^m},$$

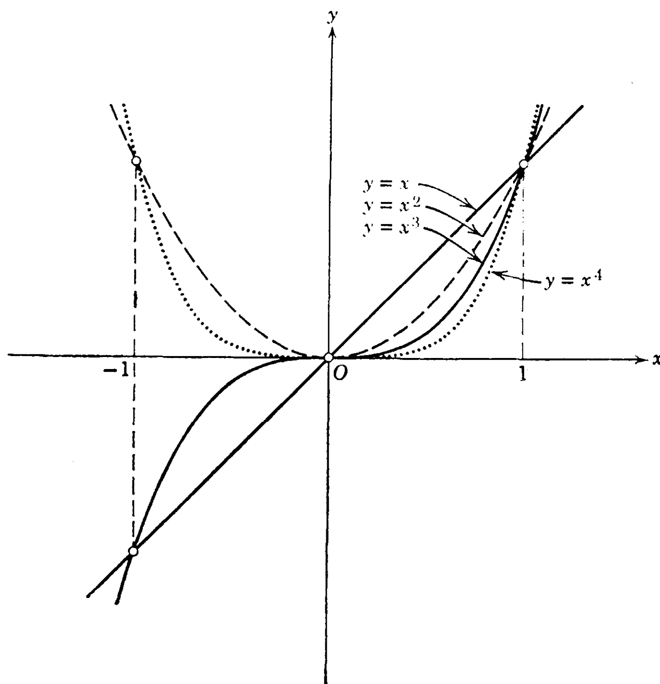


Figura 1.35 Potencias de x.

son las *funciones racionales* generales; éstas están definidas en todos los puntos donde el denominador es distinto de cero.

El polinomio más simple, la *función lineal*

$$y = ax + b,$$

se representa gráficamente por una línea recta. Toda *función cuadrática*

$$y = ax^2 + bx + c$$

se representa por una parábola. Las gráficas de polinomios de tercer grado

$$y = ax^3 + bx^2 + cx + d,$$

se llaman a veces "parábolas de tercer orden"; etc.

Las gráficas de la función $y = .x^n$ para exponentes $n = 1, 2, 3, 4$ están dadas en la Fig. 1.35. Para valores pares de n la función $y = x^n$ satisface la ecuación $f(-x) = f(x)$, y es por ello una función par; mientras que para valores impares de n la función satisface la condición $f(-x) = f(x)$, y es por lo tanto impar.

El ejemplo más simple de función racional que no es un polinomio es la función $y = 1/x$ mencionada en la p. 51; su gráfica es una hipérbola equilátera. Otro ejemplo es la función $y = 1/x^2$ (ver Fig. 1.26, p. 60).

b. Funciones algebraicas

De inmediato nos vemos obligados a salir del conjunto de funciones racionales debido al problema de formar sus inversas. El ejemplo típico es la función $\sqrt[n]{x}$, la inversa de x^n. Se ve fácilmente que la función $y = x^n$ para $x \geq 0$ es monótona creciente y continua. Tiene por ello una inversa univaluada, que se denota por el símbolo $x = \sqrt[n]{y}$, o bien, intercambiando las letras utilizadas para las variables dependientes e independientes,

$$y = \sqrt[n]{x} = x^{1/n}.$$

Por definición, esta raíz es siempre no negativa. Para valores impares de n la función x^n es monótona para todos los valores de x, incluyendo valores negativos. Consecuentemente, para valores impares de n la definición de $\sqrt[n]{x}$ puede extenderse de manera única a todos los valores de x; y en este caso $\sqrt[n]{x}$ es negativa para valores negativos de x.

Con mayor generalidad, puede considerarse la función

$$y = \sqrt[n]{R(x)},$$

donde $R(x)$ es una función racional. Otras funciones de tipo semejante se forman aplicando las operaciones racionales a una o más de estas funciones especiales. Así, por ejemplo, pueden formarse las funciones

$$y = \sqrt[m]{x} + \sqrt[n]{x^2 + 1}, \qquad y = x + \sqrt{x^2 + 1}.$$

Estas funciones son casos especiales de *funciones algebraicas.* (El concepto general de una función algebraica se definirá en el Volumen II.)

c. Funciones trigonométricas

Las funciones racionales y las funciones algebraicas se definen directamente por medio de las operaciones elementales de cálculo, pero la geometría es la fuente de la cual obtenemos primero ejemplos de otras funciones, las así llamadas *funciones trascendentes.*[1] De éstas, se considerarán aquí las *funciones trascendentes elementales,* a saber, las funciones trigonométricas, la función exponencial y la función logaritmo.

Figura 1.36 Las funciones trigonométricas.

En investigaciones analíticas los ángulos no se miden en grados, minutos y segundos, sino en *radianes.* El ángulo que se desea medir se coloca con su vértice en el centro de un círculo de radio 1 y el tamaño del ángulo se mide por la longitud del arco de circunferencia que es cortada por el ángulo.[2] Por lo tanto, un ángulo de 180° es el mismo que un ángulo de π

[1] La palabra "trascendente" no significa nada particularmente profundo o misterioso; sugiere simplemente que la definición de estas funciones trasciende las operaciones elementales de cálculos, *"quod algebrae vires transcendit".*

[2] La medida en radiantes de un ángulo puede ser definida también como el doble del *área* del sector correspondiente sobre el círculo de radio uno.

radianes (tiene medida π, en radianes). Un ángulo de 90° tiene medida, en radianes, $\pi/2$; un ángulo de 45° tiene a $\pi/4$ como su medida en radianes; un ángulo de 360° tiene medida, en radianes, 2π. Recíprocamente, un ángulo de 1 radián expresado en grados es

$$\frac{180°}{\pi}, \quad \text{o bien, aproximadamente, } 57°\ 17'\ 45''.$$

En lo sucesivo, siempre que se hable de un ángulo x, se entenderá un ángulo cuya medida en radianes es x.

Recordemos brevemente el significado de las funciones trigonométricas $\operatorname{sen} x$, $\cos x$, $\tan x$, $\cot x$.[1] Estas están representadas en la Fig. 1.36, en la cual el ángulo x se mide a partir del segmento OC (de longitud 1), con-

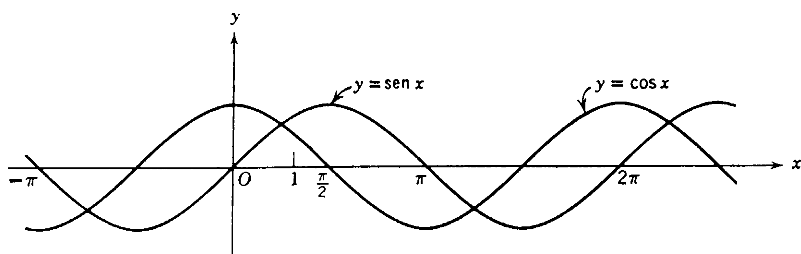


Figura 1.37.

siderando positivos a los ángulos en la dirección contraria a la de las manecillas del reloj. Las funciones $\cos x$ y $\operatorname{sen} x$ son las coordenadas rectangulares del punto A. Las gráficas de la función $\operatorname{sen} x$, $\cos x$, $\tan x$, $\cot x$ están dadas en las Figs. 1.37 y 1.38.

Más adelante (ver p. 237) estaremos en condiciones de reemplazar las definiciones geométricas por las analíticas correspondientes.

d. La función exponencial y el logaritmo

Además de las funciones trigonométricas, la función exponencial con base positiva a,

$$y = a^x,$$

y su inversa, el logaritmo de base a,

$$x = \log_a y,$$

se incluyen también entre las funciones trascendentes elementales. En matemáticas elementales es costumbre pasar por alto ciertas dificultades in-

[1] En ocasiones también es conveniente introducir las funciones $\sec x = 1/\cos x$, $\operatorname{cosec} x = 1/\operatorname{sen} x$.

herentes en su definición, y se pospondrá también aquí la discusión detallada de éstas hasta que dispongamos de mejores métodos (ver sección 2.5 p. 167). Se puede, sin embargo, indicar aquí al menos un camino "elemental" para definir estas funciones. Si $x = p/q$ es un número racional (donde p y q son enteros positivos), entonces —suponiendo positivo el número a— se define a^x como $\sqrt[q]{a^p} = a^{p/q}$, donde la raíz, de acuerdo a la convención, ha de ser tomada positiva. Puesto que los valores racionales de x son densos en todas partes, es natural extender esta función a^x a una función continua definida también para valores irracionales de x, dándole a a^x, cuando x es irracional, valores que son continuos con los valores ya definidos cuando

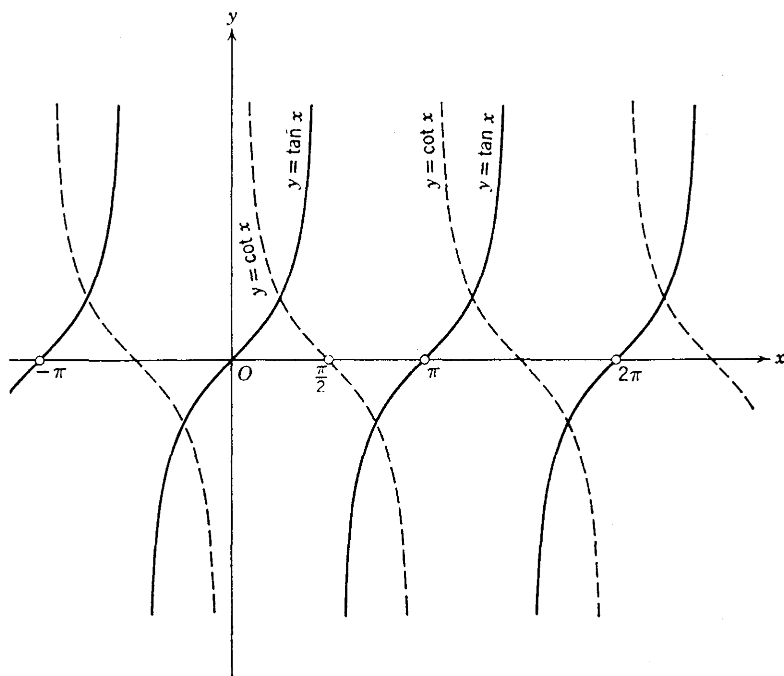


Figura 1.38.

x es irracional. Esto define una función continua $y = a^x$, la "función exponencial", la cual para todos los valores *racionales* de x dan el valor de a^x encontrado antes. Por el momento, podemos presuponer que esta extensión es realmente posible y puede ser llevada a cabo de una sola manera; pero deberá tenerse en mente que aún debemos *demostrar* que esto es cierto.[1]

La función

$$x = \log_a y$$

[1] Esto se hace en la p. 174.

puede ser entonces definida para $y > 0$ como la inversa de la función exponencial: $x = \log_a y$ es aquel número para el cual $y = a^x$.

e. Funciones compuestas, productos simbólicos, funciones inversas

Frecuentemente se forman nuevas funciones no sólo combinando funciones conocidas mediante operaciones racionales sino por el proceso más general y básico de formar *funciones de funciones* o funciones *compuestas.*

Sea $u = \phi(x)$ una función cuyo dominio está en el intervalo $a \leq x \leq b$ y cuyo rango se encuentra situado en el intervalo $\alpha \leq u \leq \beta$. Además, sea $y = g(u)$ una función definida para $\alpha \leq u \leq \beta$. Entonces $g(\phi(x)) = f(x)$ define una función f para $a \leq x \leq b$, la cual está "compuesta", o bien es la "composición" de g y ϕ. Por ejemplo, $f(x) = 1/(1 + x^{2n})$ es la composición de las funciones $\phi(x) = 1 + x^{2n}$ y $g(u) = 1/u$. Análogamente, la función $f(x) = \text{sen}\,(1/x)$ es la composición de $\phi(x) = 1/x$ y $g(u) = \text{sen}\, u$.

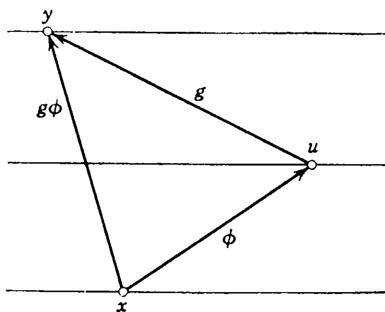


Figura 1.39 Producto simbólico $g\phi = f$ de dos funciones.

Es útil interpretar las funciones compuestas en términos de transformaciones o *"mapeos"*. La transformación ϕ lleva cada punto x del intervalo $[a, b]$ a un punto u del intervalo $[\alpha, \beta]$. La transformación g lleva cualquier valor u en $[\alpha, \beta]$ a un punto y. La transformación f es el "producto simbólico" $g\phi$ de las transformaciones g y ϕ, esto es, la formada por ϕ y g sucesivamente, en ese orden; para cualquier x en $[a, b]$ se forma su imagen u por la transformación ϕ, y entonces se aplica g a la imagen $u = \phi(x)$, obteniéndose $g(\phi(x)) = f(x) = y$ (ver Fig. 1.39). Tal producto simbólico $g\phi$ es natural y tiene sentido para cualquier tipo de operación; significa que primero se realiza ϕ, y luego sobre el resultado se aplica g.[1] No debe confun-

[1] Que el producto $g\phi$ corresponde a realizar *primero* ϕ *y luego* g (en *este* orden) parece a primera vista no natural, pero en realidad corresponde a la convención siempre adoptada en matemáticas que consiste en escribir el argumento x de una función $f(x)$ *a la derecha* del símbolo f para la función. Así, por ejemplo, en $\text{sen}(\log x)$ se sobreentiende que primero se forma el logaritmo de x y después se toma el seno de esto, y no en sentido inverso.

dirse el producto simbólico $g\phi = g(\phi)$ de dos funciones con el producto *algebraico* ordinario $g(x) \cdot \phi(x)$ de funciones, en el cual tanto $g(x)$ como $\phi(x)$ se realizan para el mismo argumento x (transformaciones aplicadas al mismo punto) y se forma el *producto de los valores* de las funciones.

Naturalmente, no puede esperarse que los productos simbólicos sean conmutativos. En general, $g(\phi)$ y $\phi(g)$ no son los mismos, aún cuando ambos estén definidos; el orden en el cual las operaciones se realizan es de la mayor importancia. Si, por ejemplo, ϕ establece la operación de "sumar 1 a un número" y g la operación de "multiplicar un número por 2", entonces

$$g(\phi(x)) = 2(x+1) = 2x+2, \qquad \phi(g(x)) = (2x)+1 = 2x+1.$$

(Ver Fig. 1.40.)

Para estar en condiciones de formar el producto simbólico $g\phi$ de dos funciones, los "factores" g y ϕ deberán adaptarse mutuamente en el sentido de que el dominio de g deberá contener al rango de ϕ. Por lo tanto, no puede formarse $g\phi$ cuando

$$g(u) = \sqrt{u}, \qquad \text{y} \qquad \phi(x) = -1 - x^2.$$

Es útil considerar funciones que están compuestas más de una vez. Una de tales funciones es

$$f(x) = \sqrt{1 + \tan(x^2)},$$

que puede construirse mediante las composiciones sucesivas

$$\phi(x) = x^2, \qquad \psi(\phi) = 1 + \tan\phi, \qquad g(\psi) = \sqrt{\psi} = f(x).$$

Podría escribirse simbólicamente $f = g\psi\phi$.

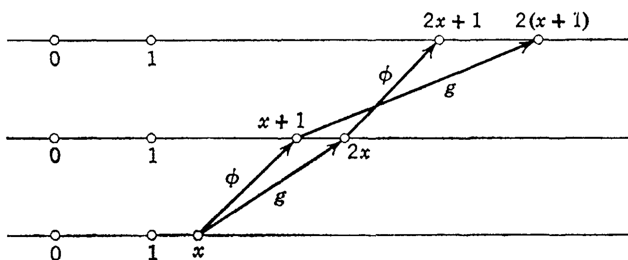


Figura 1.40 La no conmutatividad de transformaciones.

Funciones inversas

La noción de "función inversa" se aclara en el contexto del producto de transformaciones. Considérese la función o transformación ϕ que asocia a un punto x en el dominio de ϕ la imagen $u = \phi(x)$. Supóngase que la función es tal que x diferentes son transformados siempre en diferentes u,

y recíprocamente. La función se denomina entonces biunívoca. Ahora bien, un valor u es la imagen de a lo más un valor x. Puede asociarse con cada u en el rango de ϕ el valor $x = g(u)$ del cual u es la imagen bajo la transformación ϕ. De esta manera se ha definido una función g cuyo dominio es el rango de ϕ, y el cual, cuando es aplicado a una imagen $u = \phi(x)$, reproduce el valor original x; esto es, $g(\phi(x)) = x$. La función g se denomina la *inversa* de ϕ. Está caracterizada por la ecuación simbólica $g\phi x = x$.

La transformación identidad

La transformación identidad, I, se define como aquella que transforma a cada x en sí mismo. Para la inversa g de ϕ se tiene que $g\phi = I$.[1] La transformación I desempeña el mismo papel en lo que respecta a la multiplicación simbólica que el número 1 en la multiplicación ordinaria: la multiplicación por I no modifica a una transformación cualquiera dada. Por lo tanto, la ecuación $g\phi = I$ sugiere la notación $g = \phi^{-1}$ para la inversa de ϕ. Por ejemplo, la inversa $x = \operatorname{arc\,sen} u$ de la función $u = \operatorname{sen} x$ se denota a menudo por $x = \operatorname{sen}^{-1} u$.[2]

A partir de la definición de la inversa g de ϕ, se sigue inmediatamente que ϕ es también la inversa de g, de modo que no sólo $g(\phi) = x$ sino que también $\phi(g(u)) = u$.

* Una función monótona $u = \phi(x)$ definida en un intervalo $a \leq x \leq b$ evidentemente define una transformación biunívoca de ese intervalo. Si, además, ϕ es continua, entonces, como se vio anteriormente, como consecuencia del teorema del valor intermedio (p. 68), es claro que el rango de ϕ, es el intervalo con puntos extremos $\phi(a)$ y $\phi(b)$. En este caso existe la inversa g de ϕ, que es también monótona y continua en ese último intervalo. En efecto, *las funciones monótonas continuas son las únicas funciones continuas que tienen inversas o que definen transformaciones biunívocas*. Así, sea $u = \phi(x)$ una función continua en el intervalo cerrado $[a, b]$ que transforma x diferentes del intervalo en diferentes u. Entonces, en particular, los valores $\phi(a) = \alpha$ y $\phi(b) = \beta$ son distintos. Supóngase que $\alpha < \beta$. Puede entonces mostrarse que $\phi(x)$ es monótona creciente en todo el intervalo. Pues, si no fuera éste el caso, podrían encontrarse dos valores c y d tales que $a \leq c < d \leq b$, para los cuales $\phi(d) < \phi(c)$. Si se tuviera también $\phi(d) \geq \phi(a)$, se seguiría del teorema del valor intermedio que existe un ξ en el intervalo $[a, c]$ para el cual $\phi(\xi) = \phi(d)$. Este ξ sería diferente de d, y la transformación no podría ser biunívoca. Si, por otra parte, $\phi(d) < \phi(a) = \alpha$, se seguiría que $\phi(a)$ es un punto intermedio entre $\phi(d)$ y $\phi(b)$; habría entonces un ξ intermedio entre d y b para el cual $\phi(\xi) = \phi(a)$, y esto contradice la naturaleza biunívoca de la transformación.

Una propiedad importante de las funciones compuestas, casi evidente, es que $g(\phi(x)) = f(x)$ es continua (donde esté definida) si g y ϕ lo son. Así, para un ε positivo dado se tiene

$$|f(x) - f(x_0)| = |g(\phi(x)) - g(\phi(x_0))| < \varepsilon \quad \text{para} \quad |\phi(x) - \phi(x_0)| < \delta,$$

[1] Más precisamente, $g\phi$ coincide con I en el dominio de ϕ.
[2] Esto no se deberá confundir con el *recíproco algebraico* $1/(\operatorname{sen} u)$.

como consecuencia de la continuidad de la función g. Sin embargo, puesto que ϕ también es continua, ciertamente se tiene $|\phi(x) - \phi(x_0)| < \delta$ para todo x que satisface $|x - x_0| < \delta'$ con algún δ' positivo adecuado. Por tanto,

$$|f(x) - f(x_0)| < \varepsilon \qquad \text{para} \qquad |x - x_0| < \delta';$$

lo que muestra la continuidad de f.

Es mucho más fácil apelar a este teorema general al demostrar la continuidad de funciones compuestas como $\sqrt{1 - x^2}$ que tratar de construir directamente un módulo de continuidad para la función.

1.4 Sucesiones

Hasta ahora se han considerado funciones de una variable continua, o bien funciones cuyos dominios consisten de uno o más intervalos. Sin embargo, en matemáticas ocurren numerosos casos en los cuales una cantidad a depende de un entero positivo n. Tal función $a(n)$ asocia un valor a cada número natural n. La función $a(n)$ se denomina *sucesión,* más específicamente, *sucesión infinita,* si n recorre *todos* los enteros positivos. Usualmente se escribe a_n[1] en lugar de $a(n)$ para el "n-ésimo elemento" de la sucesión, y se consideran los elementos que forman una sucesión arreglados en orden de subíndices crecientes n:

$$a_1, a_2, a_3, \ldots .$$

Aquí la dependencia de los números a_n con respecto de n puede ser definida por cualquier ley que sea; y, en particular, los valores a_n no tienen que ser distintos uno del otro. La idea de sucesión será comprendida más fácilmente mediante ejemplos.

1. La suma de los primeros n enteros

$$S(n) = 1 + 2 + 3 + \cdots + n = \tfrac{1}{2}n(n + 1)$$

es una función de n, dando lugar a la sucesión

$$1, 3, 6, 10, 15, \ldots .$$

2. Otra función sencilla de n es la expresión "factorial de n", el producto de los primeros n enteros:

$$n! = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot n.$$

3. Todo entero $n > 1$ que no sea un número primo es divisible por más de dos enteros positivos, mientras que los números primos son divisibles solamente por ellos mismos y por 1. Es claro que puede considerarse el

[1] Se pronuncia "*a*-sub-*n*".

número $T(n)$ de divisores de n como una función de n. Para los primeros números está dado por la tabla:

$n =$	1	2	3	4	5	6	7	8	9	10	11	12
$T(n) =$	1	2	2	3	2	4	2	4	3	4	2	6

4. Una sucesión de gran importancia en la Teoría de Números es $\pi(n)$, el número de primos menores que el número n. Su investigación detallada es uno de los problemas más fascinantes. El resultado principal es: "El número $\pi(n)$ está dado *asintóticamente*,[1] para valores grandes de n, por la función $n/\log n$". Aquí, por $\log n$ se entiende el logaritmo de "base natural" e, que se define posteriormente (pág. 100).

1.5 Inducción matemática

Se inserta aquí una discusión de un tipo muy importante de razonamiento que está difundido en la mayor parte del pensamiento matemático.

El hecho de que toda la sucesión de números naturales sea generada partiendo con el número 1 y pasando de n a $n + 1$, da lugar al fundamental "principio de inducción matemática". En las ciencias naturales se obtiene mediante "inducción empírica" de un gran número de muestras una ley que se espera se cumpla generalmente. El grado de certidumbre de la ley depende entonces del número de veces que una muestra o un "evento" haya sido observado y así confirmada la ley. Este tipo de inducción puede ser abrumadoramente convincente, aún cuando no lleve consigo la certidumbre lógica de una demostración matemática.

La *inducción matemática* es utilizada para establecer con certidumbre lógica la exactitud de un teorema para una sucesión infinita de casos. Denótese por A un aserto que se refiere a un número natural arbitrario n. Por ejemplo, la propiedad A puede ser "La suma de los ángulos interiores en un polígono simple de $n + 2$ lados es n veces 180°", o sea, $n\pi$. Para demostrar una propiedad de este tipo no es suficiente probarla para los primeros 10 o los primeros 100, o aún los primeros 1,000 valores de n. En lugar de eso debe aplicarse un método matemático que se explica primero para este ejemplo. Para $n = 1$ el polígono se reduce a un triángulo, para el cual es sabido que la suma de los ángulos es de 180°. Para un cuadrángulo, correspondiente a $n = 2$, trazamos una diagonal dividiendo el cuadrángulo en dos triángulos. Esto muestra que la suma de los ángulos del cuadrángulo es igual a la suma combinada de los ángulos de los dos triángulos, esto es, $180° + 180° = 2 \cdot 180°$. Procediendo al ejemplo de un pentágono, éste puede dividirse en un cuadrángulo y un triángulo mediante

[1] Esto es, el cociente del número $\pi(n)$ por el número $n/\log n$ difiere arbitrariamente poco de uno, bastando sólo con que n sea suficientemente grande.

el trazo de una diagonal adecuada. Esto da para la suma de los ángulos del pentágono el valor $2 \cdot 180° + 1 \cdot 180° = 3 \cdot 180°$. Puede continuarse de esta manera y probarse el teorema general sucesivamente para $n = 4$, 5, etcétera. La exactitud de la propiedad A para cualquier n se sigue de su exactitud para el *n precedente;* de esta manera se establece su validez general para todo n.

Formulación general

Lo que es esencial en la demostración de la propiedad A del ejemplo es que A se ha probado sucesivamente para los casos especiales A_1, A_2, ... A_n, ... La posibilidad de hacer esto depende de dos factores: (1) se ha de dar una demostración general mostrando que el aserto A_{r+1} es correcto siempre que A_r sea correcto; y (2) el aserto A_1 debe ser probado. Que estas condiciones son suficientes para probar la exactitud de todo aserto A_1, A_2, A_3, ..., constituye el *principio de inducción matemática.* En lo que sigue se acepta la validez de este principio como un hecho básico de lógica.

El principio puede ser formulado en una forma abstracta más general. "Sea S cualquier conjunto que consista de números naturales, que tenga las siguientes dos propiedades: (1) siempre que S contenga un número r, entonces contendrá también al número $r + 1$; y (2) S contiene al número 1. Entonces es cierto que S es el conjunto de todos los números naturales." La formulación previa del principio de inducción matemática se obtiene fácilmente si se toma como S el conjunto de los números naturales para los cuales la aseveración A es correcta.

El principio se aplica a menudo sin hacer mención específica del mismo; o bien, su uso se indica solamente mediante la expresión "etc.". Esto sucede a menudo, particularmente en matemáticas elementales. Sin embargo, en situaciones más complicadas es preferible apelar explícitamente al principio.

Ejemplos. Como ilustración se dan dos aplicaciones.

Primero se prueba una fórmula para la suma de los n primeros cuadrados. Se encuentra para n pequeños (digamos $n < 5$) que la siguiente fórmula,[1] denotada por A_n, se cumple:

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

Se conjetura que esta fórmula es correcta para todo n. Para demostrarlo, se supone que r es cualquier número para el cual la fórmula A_r es correcta, esto es, que

$$1^2 + 2^2 + 3^2 + \cdots + r^2 = \frac{r(r+1)(2r+1)}{6};$$

sumando $(r+1)^2$ a ambos miembros se obtiene

[1] Este resultado fue utilizado por el matemático griego Arquímedes en su trabajo sobre espirales.

$$1^2 + 2^2 + \cdots + r^2 + (r+1)^2 = \frac{r(r+1)(2r+1)}{6} + (r+1)^2$$

$$= \frac{(r+1)(r+2)[2(r+1)+1]}{6}$$

Esto, sin embargo, es precisamente la propiedad A_{r+1} obtenida substituyendo $r + 1$ por n en A_n. Por lo tanto, la veracidad de A_r implica la de A_{r+1}. Para completar la demostración de A_n para n general se requiere verificar solamente la exactitud de A_1, esto es, de

$$1^2 = \frac{1 \cdot 2 \cdot 3}{6}.$$

Puesto que esto es obviamente correcto, la fórmula A_n queda establecida para todo número natural n.

El lector deberá probar mediante un argumento análogo que

$$1^3 + 2^3 + 3^3 + \cdots + n^3 = \left[\frac{n(n+1)}{2}\right]^2.$$

Como ilustración adicional del principio de inducción se prueba

El teorema del binomio. *El aserto A_n del teorema está representado por la fórmula*

$$(a+b)^n = a^n + \frac{n}{1}a^{n-1}b + \frac{n(n-1)}{1 \cdot 2}a^{n-2}b^2$$

$$+ \frac{n(n-1)(n-2)}{1 \cdot 2 \cdot 3}a^{n-3}b^3 + \cdots + \frac{n(n-1)(n-2)\cdot \cdots \cdot 2 \cdot 1}{1 \cdot 2 \cdot 3 \cdot \cdots \cdot (n-1) \cdot n}b^n.$$

Es costumbre escribir esta fórmula en la forma

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{n}b^n$$

donde el *coeficiente binomial* $\binom{n}{k}$ está definido por

$$\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}$$

para $k = 1, 2, \ldots, (n-1)$, y

$$\binom{n}{0} = \binom{n}{n} = 1.$$

(Si se define $0! = 1$, la fórmula general para $\binom{n}{k}$ se aplica también a los casos $k = 0$ y $k = n$.)

Si A_n se cumple para un cierto n, se encuentra, multiplicando ambos miembros por $(a + b)$, que

$$\begin{aligned}(a+b)^{n+1} &= (a+b)\left[\binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \cdots + \binom{n}{n}b^n\right] \\ &= \binom{n}{0}a^{n+1} + \left[\binom{n}{0} + \binom{n}{1}\right]a^n b + \left[\binom{n}{1} + \binom{n}{2}\right]a^{n-1}b^2 \\ &\quad + \cdots + \left[\binom{n}{n-1} + \binom{n}{n}\right]ab^n + \binom{n}{n}b^{n+1}.\end{aligned}$$

Ahora bien,

$$\begin{aligned}&\binom{n}{k} + \binom{n}{k+1} \\ &\quad = \frac{n(n-1)\cdots(n-k+1)}{k!} + \frac{n(n-1)\cdots(n-k+1)(n-k)}{(k+1)!} \\ &\quad = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}\left(1 + \frac{n-k}{k+1}\right) \\ &\quad = \frac{(n+1)n(n-1)\cdots(n-k+1)}{(k+1)!} = \binom{n+1}{k+1}.\end{aligned}$$

Puesto que $\binom{n}{0} = \binom{n+1}{0} = 1$ y $\binom{n}{n} = \binom{n+1}{n+1} = 1$, se tiene

$$\begin{aligned}(a+b)^{n+1} &= \binom{n+1}{0}a^{n+1} + \binom{n+1}{1}a^n b + \binom{n+1}{2}a^{n-1}b^2 + \cdots \\ &\quad + \binom{n+1}{n}ab^n + \binom{n+1}{n+1}b^{n+1},\end{aligned}$$

que es la fórmula A_{n+1}. Puesto que también para $n = 1$ la fórmula vale; es decir,

$$(a+b)^1 = \binom{1}{0}a + \binom{1}{1}b = a + b,$$

resulta que el Teorema del Binomio se cumple para todos los números naturales n.

1.6 El límite de una sucesión

El concepto fundamental sobre el cual descansa la esencia toda del análisis matemático es el de *límite* de una sucesión infinita a_n. A menudo un número a se describe por medio de una sucesión infinita a_n de aproximaciones; esto es, el valor a está dado por el valor a_n con cualquier grado de precisión deseado si el índice n se elige suficientemente grande. Se han encontrado ya tales *representaciones de números a como "límites"* de sucesiones en su representación como fracciones decimales infinitas; los números reales aparecen entonces como límites, para n crecientes, de las sucesiones de fracciones decimales ordinarias con n dígitos. En la sección 1.7 se dará una discusión precisa general del concepto de límite; por el momento se ilustrará la idea de límite mediante algunos ejemplos significativos.

Las sucesiones $a_1, a_2, \ldots$ pueden ser convenientemente representadas por una sucesión de "bloques", correspondiendo el elemento a_n al rectángulo en el plano xy acotado por las rectas $x = n - 1$, $x = n$, $y = a_n$, $y = 0$, que tiene por área [1] el número $|a_n|$; o bien, equivalentemente, por la gráfica de una función constante por pedazos, $a(x)$, de una variable continua x con discontinuidades de salto en los puntos $x = n$.

a. $a_n = \dfrac{1}{n}$

Considérese la sucesión

$$1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots.$$

(Ver Fig. 1.41.) Ningún número de esta sucesión es cero; pero, conforme el número n crece, a_n se aproxima al cero. Además, si se toma cualquier

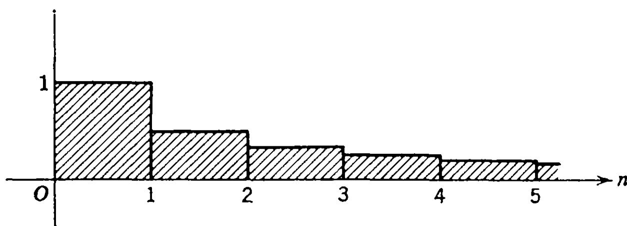


Figura 1.41 La sucesión $a_n = \dfrac{1}{n}$.

intervalo centrado en el origen, no importa cuán pequeño, entonces a partir de un índice definido hacia adelante todos los números a_n se encontrarán en ese intervalo. Esta situación se expresa diciendo que "si n crece los

[1] Pudo haberse escogido igualmente el rectángulo acotado por las rectas $x = n$, $x = n + 1$, $y = a_n$, $y = 0$ para representar a_n.

números a_n *tienden* a cero", o que "poseen el *límite* cero", o bien que "la sucesión $a_1, a_2, a_3, \ldots$ *converge* a cero".

Si los números son representados como puntos en una línea, esto significa que los puntos $1/n$ se apiñan cada vez más cerca del punto cero conforme n crece.

La situación es semejante para la sucesión

$$1, -\frac{1}{2}, \frac{1}{3}, -\frac{1}{4}, \ldots, \frac{(-1)^{n-1}}{n}, \ldots.$$

(Ver Fig. 1.42.) También aquí, los números a_n tienden a cero cuando n crece. La única diferencia es que los números a_n son a veces mayores y a veces menores que el límite cero; como se dice, la sucesión *oscila* en torno al límite.

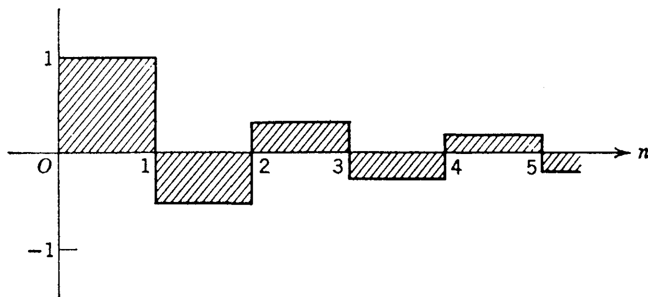


Figura 1.42 La sucesión $a_n = \frac{(-1)^{n-1}}{n}$.

La convergencia de la sucesión a cero usualmente es expresada simbólicamente por la ecuación

$$\lim_{n\to\infty} a_n = 0,$$

o bien, ocasionalmente, por la notación abreviada

$$a_n \to 0.$$

b. $a_{2m} = \dfrac{1}{m}$; $a_{2m-1} = \dfrac{1}{2m}$

En los ejemplos anteriores el valor absoluto de la diferencia entre a_n y el límite se tornaba progresivamente menor conforme n crecía. Esto no es necesariamente siempre el caso, como se muestra en la sucesión

$$\frac{1}{2}, 1, \frac{1}{4}, \frac{1}{2}, \frac{1}{6}, \frac{1}{3}, \ldots, \frac{1}{2m}, \frac{1}{m}, \ldots;$$

(ver Fig. 1.43) dada para valores pares $n = 2m$ por $a_n = a_{2m} = 1/m$; y para valores impares $n = 2m - 1$ por $a_n = a_{2m-1} = 1/2m$. Esta sucesión tiene también el límite cero; pues todo intervalo en torno al origen, por pequeño que sea, contiene a todos los números a_n desde un cierto valor

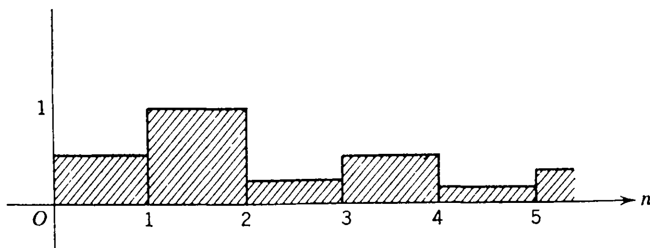


Figura 1.43 La sucesión $a_{2n} = \frac{1}{n}, a_{2n-1} = \frac{1}{2n}$.

de n hacia adelante; pero no es cierto que todo número a_n esté situado más cerca del límite cero que el número precedente.

c. $a_n = \dfrac{n}{n+1}$

Considérese la sucesión

$$a_1 = \frac{1}{2}, a_2 = \frac{2}{3}, \ldots, a_n = \frac{n}{n+1}, \ldots.$$

Escribiendo $a_n = 1 - 1/(n+1)$, se ve que al crecer n el número a_n se aproxima al número 1, en el sentido de que para todo intervalo definido en torno del punto 1 todos los números a_n sucesores a cierto a_N deberán encontrarse en ese intervalo. Se escribe

$$\lim_{n\to\infty} a_n = 1.$$

La sucesión

$$a_n = \frac{n^2 - 1}{n^2 + n + 1}$$

se comporta de una manera semejante. Esta sucesión tiende también a un límite al crecer n; de hecho, al límite uno: $\lim_{n\to\infty} a_n = 1$. Esto se ve más fácilmente si se escribe

$$a_n = 1 - \frac{n+2}{n^2 + n + 1} = 1 - r_n;$$

y se requiere mostrar solamente que los números r_n tienden a cero conforme n crece. Para todos los valores de n mayores que 2 se tiene $n + 2 < 2n$ y $n^2 + n + 1 > n^2$. Por lo tanto, para el residuo r_n se tiene

$$0 < r_n < \frac{2n}{n^2} = \frac{2}{n} \qquad (n > 2),$$

de donde se ve que r_n tiende a cero cuando n crece. Esta discusión da al mismo tiempo una estimación del máximo valor según el cual el número a_n (para $n > 2$) puede diferir del límite uno; esta diferencia no puede exceder a $2/n$.

Este ejemplo ilustra el hecho de que para valores grandes de n los términos con los mayores exponentes en el numerador y denominador de la fracción para a_n predominan y determinan el límite.

d. $a_n = \sqrt[n]{p}$

Sea p cualquier número positivo fijo. Considérese la sucesión a_1, a_2, a_3, . . . , a_n, . . . , donde

$$a_n = \sqrt[n]{p}.$$

Se afirma que

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \sqrt[n]{p} = 1.$$

Se probará esto utilizando un lema que también será útil para otros propósitos.

LEMA. *Si h es un número positivo y n un entero positivo, entonces*

(1) $$(1 + h)^n \geq 1 + nh.$$

Esta desigualdad es una consecuencia trivial del teorema del binomio (ver p. 82), de acuerdo con el cual

$$(1 + h)^n = 1 + nh + \frac{n(n-1)}{2} h^2 + \cdots + h^n,$$

si se observa que todos los términos en el desarrollo de $(1 + h)^n$ son no negativos. El mismo argumento conduce a la desigualdad más fuerte

$$(1 + h)^n \geq 1 + nh + \frac{n(n-1)}{2} h^2.$$

Volviendo a la sucesión, se distinguirá entre los casos $p > 1$ y $p < 1$ (si $p = 1$, entonces $\sqrt[n]{p}$ es igual a 1 para todo n, y el aserto es ciertamente verdadero).

Si $p > 1$, entonces $\sqrt[n]{p}$ es también mayor que 1; hágase $\sqrt[n]{p} = 1 + h_n$, donde h_n es una cantidad positiva dependiente de n; y por la desigualdad (1) se tiene

$$p = (1 + h_n)^n \geq 1 + nh_n,$$

implicando

$$0 < h_n \leq \frac{p-1}{n}.$$

Cuando n crece el número h_n deberá tender a cero, lo que prueba que a_n converge al límite uno, como se afirmó. Al mismo tiempo se dispone de un medio para estimar cuán cerca está a_n del límite uno, puesto que la diferencia h_n entre a_n y uno no es mayor que $(p - 1)/n$.

Si $p < 1$, entonces $1/p > 1$, y $\sqrt[n]{1/p}$ converge al límite uno. Sin embargo,

$$\sqrt[n]{p} = \frac{1}{\sqrt[n]{1/p}}.$$

Como el recíproco de una cantidad que tiende a uno, el propio $\sqrt[n]{p}$ tiende a uno.

e. $a_n = \alpha^n$

Considérese la sucesión $a_n = \alpha^n$, donde α es fijo y n recorre la sucesión de enteros positivos.

Primero, sea α un número positivo menor que uno. Hágase entonces $\alpha = 1/(1 + h)$, donde h es positivo, y la desigualdad (1) da

$$a_n = \frac{1}{(1 + h)^n} \leq \frac{1}{1 + nh} < \frac{1}{nh}$$

Puesto que h, y consecuentemente $1/h$, depende solamente de α y no cambia cuando n crece, se ve que α^n tiende a cero conforme n crece:

$$\lim_{n \to \infty} \alpha^n = 0 \qquad (0 < \alpha < 1).$$

Se satisface la misma relación cuando α es cero, o bien negativo pero mayor que -1. Esto es inmediatamente obvio, puesto que en cualquier caso $\lim_{n \to \infty} |\alpha|^n = 0$.

Si $\alpha = 1$, entonces α^n es siempre igual a uno y se tendrá que considerar al número uno como el límite de α^n.

Si $\alpha > 1$, hágase $\alpha = 1 + h$, donde h es positivo, y se verá de inmediato, de la desigualdad, que cuando n crece α^n no tiende hacia algún límite definido sino que crece más allá de toda cota. Se dice que α^n *tiende a infinito* cuando n crece indefinidamente, o bien que α^n *se torna infinito;* en símbolos,

$$\lim_{n \to \infty} \alpha^n = \infty \qquad (\alpha > 1).$$

Se insiste explícitamente en que *el símbolo* ∞ *no denota un número y que no puede calcularse con él de acuerdo a las reglas usuales;* asertos que afirmen que una cantidad es o se vuelve infinita jamás tendrán el mismo sentido que una aseveración que involucre cantidades definidas finitas. A pesar de esto, tales modos de expresión y el empleo del símbolo ∞ son extremadamente convenientes, como a menudo se verá en las siguientes páginas.

Si $\alpha = -1$, el valor α^n no tiende hacia límite alguno, pero cuando n crece y recorre la sucesión de enteros positivos α^n toma los valores $+1$ y -1 alternativamente. Análogamente, si $\alpha < -1$ el valor α^n crece numéricamente más allá de toda cota, pero el signo es alternativamente positivo y negativo.

f. Ilustración geométrica de los límites α^n y $\sqrt[n]{p}$

Si consideramos las gráficas de las funciones $y = x^n$ y $y = x^{1/n} = \sqrt[n]{x}$ y, por razones de conveniencia, nos restringimos a valores no negativos

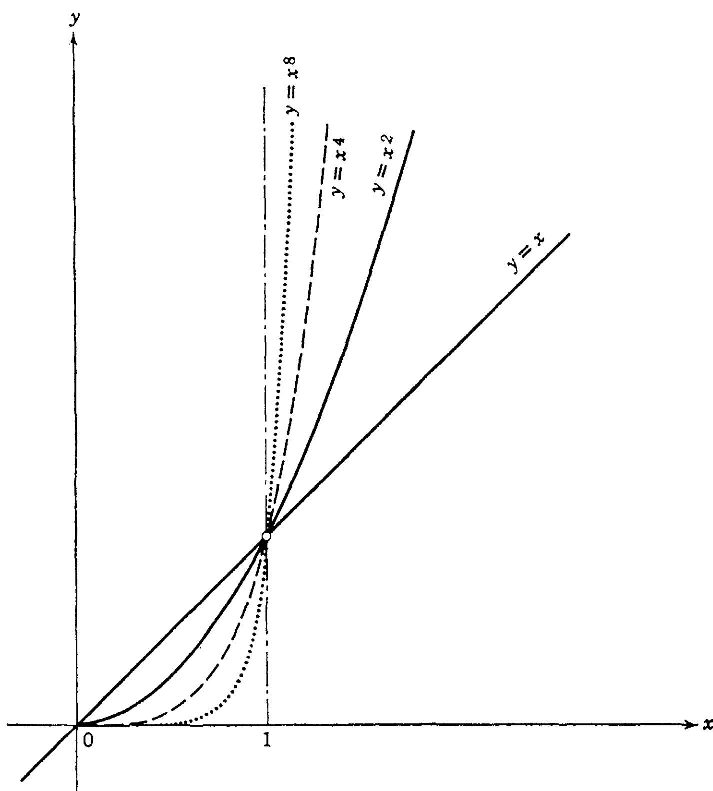


Figura 1.44 x^n conforme n crece.

de x, los límites anteriores están ilustrados en las Figs. 1.44 y 1.45, respectivamente. Se observa que en el intervalo de 0 a 1 las curvas $y = x^n$ se acercan más y más al eje x cuando n crece, mientras que fuera del intervalo se elevan cada vez con mayor pendiente y se aproximan a una

línea paralela al eje y. Todas las curvas pasan por el punto con coordenadas $x = 1$, $y = 1$ y por el origen.

Las gráficas de las funciones $y = x^{1/n} = + \sqrt[n]{x}$, se acercan cada vez más a la línea paralela al eje x y a una distancia 1 por encima de éste; y aquí también todas las curvas deberán pasar por el origen y el punto $(1, 1)$. De aquí que en el límite las curvas aproximan la línea quebrada que consiste de la parte del eje y entre los puntos $y = 0$ e $y = 1$ y de la paralela al eje x, $y = 1$. Además, es evidente que las dos figuras se encuentran muy relacionadas, como podría esperarse del hecho de que las funciones $y = \sqrt[n]{x}$ son las funciones inversas de las n-ésimas potencias, de lo cual se infiere de que para cada n la gráfica de $y = x^n$ es transformada en la de $y = \sqrt[n]{x}$ mediante una reflexión con respecto de la recta $y = x$.

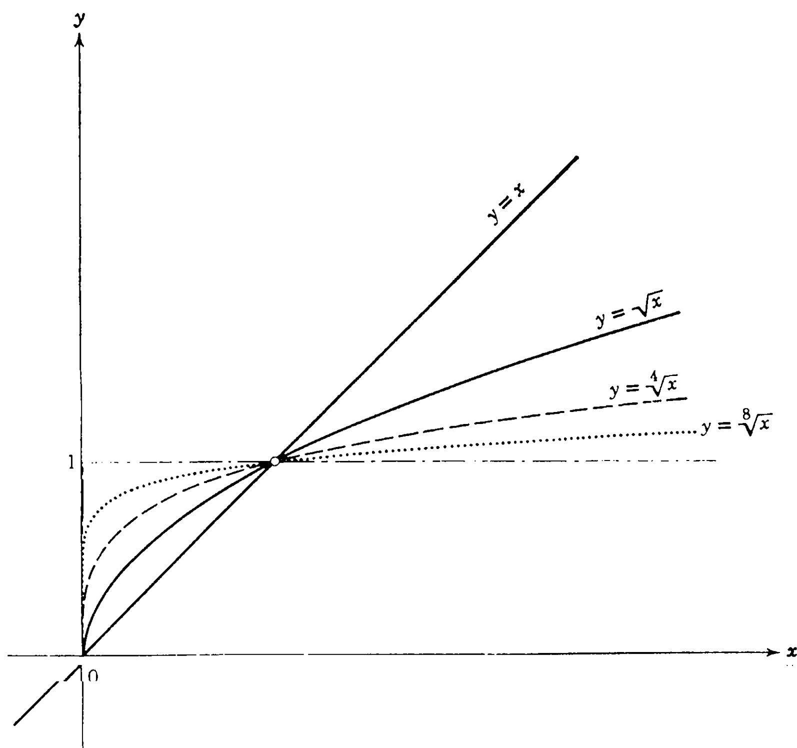


Figura 1.45 $x^{1/n}$ conforme n crece.

g. La serie geométrica

Un ejemplo de un límite muy conocido en las matemáticas elementales es proporcionado por la *serie geométrica:*

$$1 + q + q^2 + \cdots + q^{n-1} = S_n;$$

El número q se denomina la *razón común* o *cociente* de la serie. El valor de esta suma puede, como es bien sabido, expresarse en la forma

$$S_n = \frac{1 - q^n}{1 - q}$$

siempre que $q \neq 1$; y puede obtenerse esta expresión multiplicando la suma S_n por q y restando la ecuación así obtenida de la ecuación original, o bien puede verificarse la fórmula por división.

¿Qué sucede a la suma S_n cuando n crece indefinidamente? La respuesta es la siguiente: La sucesión de sumas S_n tiene un límite definido S si q está situado entre -1 y $+1$, excluyendo estos valores extremos, y

$$S = \lim_{n \to \infty} S_n = \frac{i}{1 - q}.$$

Para verificar esta propiedad se escribe S_n como $(1 - q^n)/(1 - q) = 1/(1 - q) - q^n/(1 - q)$. Se ha probado ya que siempre que $|q| < 1$ la cantidad q^n tiende a cero si n crece; de aquí que, bajo esta suposición, $q^n/(1 - q)$ tiende también a cero y S_n tiende al límite $1/(1 - q)$ cuando n crece.

El paso al límite, $\lim_{n \to \infty} (1 + q + q^2 + \cdots + q^{n-1}) = 1/(1 - q)$, se expresa usualmente diciendo que cuando $|q| < 1$ *la suma de la serie geométrica infinita es la expresión* $1/(1 - q)$.

Las sumas S_n de la serie geométrica finita también se denominan *sumas parciales* de la serie geométrica infinita $1 + q + q^2 + \ldots$ (Debe hacerse una distinción entre la *sucesión* de números q^n y las sumas parciales de la *serie* geométrica.)

El hecho de que las sumas parciales S_n de la serie geométrica tiendan al límite $S = 1/(1 - q)$ conforme n crece se expresa también diciendo que la serie geométrica infinita $1 + q + q^2 + \ldots$ *converge* a la suma $S = 1/(1 - q)$ cuando $|q| < 1$.

De paso, deberá notarse que si q es racional, por ejemplo, $q = \frac{1}{2}$, o bien $q = \frac{1}{3}$, entonces la suma de la serie geométrica infinita tiene un valor racional (en los casos mencionados los valores son 2 y $\frac{3}{2}$, respectivamente). Esta observación está detrás del hecho bien conocido de que las *fracciones decimales periódicas representan siempre números racionales.*[1] La demostración general de esta propiedad resultará evidente del ejemplo del número

$$x = 0.343434 \cdots,$$

que puede ser evaluado escribiendo

[1] Véase Courant y Robbins, *What Is Mathematics?*, p. 66.

$$\begin{aligned} x &= \frac{34}{10^2} + \frac{34}{10^4} + \frac{34}{10^6} + \cdots \\ &= \frac{34}{10^2}\left(1 + \frac{1}{10^2} + \frac{1}{10^4} + \cdots\right) \\ &= \frac{34}{100}\frac{1}{1 - 1/100} = \frac{34}{99}. \end{aligned}$$

h. $a_n = \sqrt[n]{n}$

Mostraremos que la sucesión de números

$$a_1 = 1, \qquad a_2 = \sqrt{2}, \qquad a_3 = \sqrt[3]{3}, \ldots, \qquad a_n = \sqrt[n]{n}, \ldots$$

tiende a 1 cuando n crece:

$$\lim_{n\to\infty} \sqrt[n]{n} = 1.$$

Puesto que a_n excede el valor 1, hacemos $a_n = 1 + h_n$, con h_n positivo. Entonces (ver p. 87)

$$\begin{aligned} n &= (a_n)^n = (1 + h_n)^n \\ &\geq 1 + nh_n + \frac{n(n-1)}{2} h_n^2 \geq \frac{n(n-1)}{2} h_n^2. \end{aligned}$$

Se sigue para $n > 1$ que

$$h_n^2 \leq \frac{2}{n-1};$$

y de aquí que

$$h_n \leq \frac{\sqrt{2}}{\sqrt{n-1}}.$$

Se tiene ahora

$$1 \leq a_n = 1 + h_n \leq 1 + \frac{\sqrt{2}}{\sqrt{n-1}}.$$

El miembro derecho de esta desigualdad obviamente tiende a uno, y por lo tanto también a_n.

i. $a_n = \sqrt{n+1} - \sqrt{n}$

En este ejemplo los a_n son diferencias de dos términos, cada uno de los cuales crece más allá de toda cota. Intentando pasar al límite separadamente con cada uno de los dos términos, se obtiene la expresión simbólica sin significado $\infty - \infty$. En tal caso la existencia de un límite, y lo que pueda ser su valor, depende totalmente del caso especial. Se afirma que en este ejemplo es

$$\lim_{n\to\infty} (\sqrt{n+1} - \sqrt{n}) = 0.$$

Para la demostración se requiere escribir solamente la expresión en la forma

$$\sqrt{n+1} - \sqrt{n} = \frac{(\sqrt{n+1} - \sqrt{n})(\sqrt{n+1} + \sqrt{n})}{\sqrt{n+1} + \sqrt{n}} = \frac{1}{\sqrt{n+1} + \sqrt{n}};$$

y ver de inmediato que tiende a cero conforme n crece.

j. $a_n = \frac{n}{\alpha^n}$, **para** $\alpha > 1$

Formalmente, el límite de los a_n es del tipo indeterminado ∞/∞ encontrado ya en el ejemplo *c*. Se afirma que en este ejemplo la sucesión de números $a_n = n/\alpha^n$ tiende al límite cero.

Para la demostración hágase $\alpha = 1 + h$, donde $h > 0$, y hágase uso nuevamente de la desigualdad

$$(1 + h)^n \geq 1 + nh + \frac{n(n-1)}{2} h^2 > \frac{n(n-1)}{2} h^2.$$

De aquí que para $n > 1$

$$a_n = \frac{n}{(1+h)^n} < \frac{2}{(n-1)h^2}.$$

Puesto que a_n es positivo y el miembro derecho de esta desigualdad tiende a cero, a_n deberá tender también a cero.

1.7 Discusión del concepto de límite

a. Definición de convergencia y divergencia

De los ejemplos discutidos en la sección 1.6 puede abstraerse el siguiente concepto general de límite:

Supóngase que para una dada sucesión infinita de puntos $a_1, a_2, a_3, \ldots,$ existe un número finito 1 tal que todo intervalo abierto, por pequeño que sea, definido en torno del punto 1, contiene todos los puntos a_n excepto, a lo más, un número finito. El número 1 se denomina entonces límite de la sucesión $a_1, a_2, \ldots$; o bien se dice que la sucesión $a_1, a_2, \ldots$ es convergente y converge a 1. En símbolos $\lim_{n\to\infty} a_n = l$.

La siguiente definición de límite es equivalente:

A todo número positivo ε, por pequeño que sea, podemos asignarle un entero $N = N(\varepsilon)$ suficientemente grande tal que a partir del índice N en adelante [esto es, para $n > N(\varepsilon)$] se tiene siempre $|a_n - l| < \varepsilon$.

Por supuesto, es cierto como regla que $N(\varepsilon)$ deberá ser escogido cada vez mayor para valores cada vez menores de la tolerancia o error ε; en otras palabras, $N(\varepsilon)$ usualmente crecerá más allá de toda cota a medida que ε tienda a cero. La noción intuitivamente vaga de límite sugiere una representación de los a_n *moviéndose cada vez más hacia 1.* Esta representación es reemplazada aquí por la definición "estática" precisa: Toda vecindad de l contiene a todos los a_n excepto, a lo más, un número finito de ellos.[1]

Obviamente, una sucesión $a_1, a_2, \ldots$, no puede tener más de un límite l. Si, por el contrario, dos números distintos l y l' fueran límites de la misma sucesión $a_1, a_2, \ldots$, podrían determinarse intervalos abiertos en torno de cada uno de los puntos l y l' que no se traslapen. Puesto que cada intervalo contiene todos los a_n excepto un número finito de ellos, la sucesión no podría ser infinita. El límite de una sucesión convergente está por ello determinado unívocamente.

Otra observación evidente pero útil es la siguiente: Si de una sucesión convergente se omite cualquier número de términos, la sucesión que resulta converge al mismo límite que la sucesión original.

Una sucesión que no converge se dice que es divergente. Si a medida que n crece los números a_n crecen más allá de toda cota positiva, se dice que la sucesión *diverge* a $+\infty$; como se ha hecho ya ocasionalmente, se escribe entonces $\lim_{n\to\infty} a_n = \infty$. De manera análoga, se escribe $\lim_{n\to\infty} a_n = -\infty$ si, cuando n crece, los números $-a_n$ crecen más allá de toda cota en la dirección positiva. Sin embargo, la divergencia puede manifestarse por sí misma de otras maneras, tal como para la sucesión $a_1 = -1$, $a_2 = +1$, $a_3 = -1$, $a_4 = +1, \ldots$, cuyos términos oscilan de un sitio a otro entre valores diferentes.

Evidentemente, ni la divergencia ni la convergencia de una sucesión son afectadas si se omite un número finito de términos.

Una sucesión $a_1, a_2, \ldots$ es *acotada* si existe un intervalo finito que contiene *todos* los puntos de la sucesión. Cualquier intervalo finito está contenido en algún intervalo finito que tiene el origen como centro. Por lo tanto, el requisito de que la sucesión sea acotada significa que existe un número M tal que $|a_n| \leq M$ para todo n.

Una sucesión convergente $a_1, a_2, \ldots$ también es necesariamente acotada. En efecto, sea l el límite de la sucesión. Tomando $\varepsilon = 1$ se encuentra a partir de la definición de convergencia que toda a_n a partir de cierta N en adelante está en el intervalo de longitud 2 con centro en l. Los únicos términos a_n de la sucesión que pueden estar fuera de ese intervalo son $a_1, \ldots, a_{N-1}$. Puede entonces, sin embargo, encontrarse un intervalo finito mayor que incluya también a $a_1, \ldots, a_{N-1}$.

---

[1] El lector notará la analogía con la definición de continuidad de una función $f(x)$ en un punto x_0. El papel que la cantidad suficientemente pequeña $\delta(\varepsilon)$ desempeña ahí es desempeñada aquí por el entero suficientemente grande $N(\varepsilon)$. En realidad, se verá en la p. 105 que la continuidad de una función es un punto puede ser formulada en términos de límites de sucesiones.

b. Operaciones racionales con límites

De la definición de límite se sigue de inmediato que se pueden efectuar las operaciones elementales de adición, multiplicación, substracción y división de límites de acuerdo con las siguientes reglas.

Si $a_1, a_2, \ldots$ es una sucesión con límite a y $b_1, b_2, \ldots$ es una sucesión con límite b, entonces la sucesión de números $c_n = a_n + b_n$ tiene también un límite c, y

$$c = \lim_{n\to\infty} c_n = a + b.$$

La sucesión de números $c_n = a_n b_n$ converge de la misma manera y

$$\lim_{n\to\infty} c_n = ab.$$

Análogamente, la sucesión $c_n = a_n - b_n$ converge y

$$\lim_{n\to\infty} c_n = a - b.$$

Siempre que el límite b difiera de cero, los números $c_n = a_n/b_n$ convergen de manera análoga y tienen el límite

$$\lim_{n\to\infty} c_n = \frac{a}{b}.$$

En palabras: Podemos *intercambiar* las operaciones racionales de cálculo con el proceso de formar límite; se obtiene el mismo resultado ya sea que se efectúe primero el paso al límite y después una operación racional o viceversa.

Las demostraciones de todas estas reglas resultan claras si una de ellas es llevada a cabo. Considérese la multiplicación de límites. Si las relaciones $a_n \to a$ y $b_n \to b$ valen, entonces para todo número positivo ε podemos asegurar ambas desigualdades

$$|a - a_n| < \varepsilon \qquad \text{y} \qquad |b - b_n| < \varepsilon$$

escogiendo n suficientemente grande, digamos $n > N(\varepsilon)$. Si se escribe

$$ab - a_n b_n = b(a - a_n) + a_n(b - b_n)$$

y se recuerda que existe una cota positiva M independiente de n tal que $|a_n| < M$, se obtiene

$$|ab - a_n b_n| \leq |b|\,|a - a_n| + |a_n|\,|b - b_n| < (|b| + M)\varepsilon.$$

Puesto que la cantidad $(|b| + M)\varepsilon$ puede hacerse arbitrariamente pequeña escogiendo ε suficientemente pequeño, la diferencia entre ab y $a_n b_n$ se hace en efecto tan pequeña como se desee para valores suficientemente grandes de n; y ésta es precisamente la aseveración expresada en la ecuación

$$ab = \lim_{n\to\infty} a_n b_n.$$

Usando este ejemplo como modelo, el lector puede demostrar las reglas para las operaciones racionales restantes.

Por medio de estas reglas muchos límites pueden ser evaluados fácilmente; en efecto se tiene:

$$\lim_{n\to\infty} \frac{n^2 - 1}{n^2 + n + 1} = \lim_{n\to\infty} \frac{1 - \dfrac{1}{n^2}}{1 + \dfrac{1}{n} + \dfrac{1}{n^2}} = 1,$$

puesto que en la segunda expresión puede pasarse directamente al límite en el numerador y en el denominador.

La siguiente regla simple es frecuentemente útil: *Si* $\lim a_n = a$ *y* $\lim b_n = b$, *y si, además,* $a_n > b_n$ *para todo* n, *entonces* $a \geq b$. Sin embargo, de ningún modo hay que esperar que a sea siempre *mayor* que b, como lo muestran las sucesiones $a_n = 1/n$, $b_n = 1/2n$, para las cuales $a = b = 0$.

c. Pruebas intrínsecas de convergencia. Sucesiones monótonas

En todos los ejemplos dados el límite de la sucesión considerada era un número conocido De hecho, para aplicar la definición anterior de límite de una sucesión debe conocerse el límite antes de que pueda verificarse la convergencia. Si del concepto de límite de una sucesión no se obtuviera otra cosa más que el reconocimiento de que algunos números conocidos pueden ser aproximados por ciertas sucesiones de otros números conocidos, se hubiera ganado muy poco con ello. La ventaja del concepto de límite en análisis radica esencialmente en el hecho de que problemas importantes tienen a menudo soluciones numéricas que posiblemente de otra forma no podrían ser conocidas o expresables directamente, pero que pueden ser descritas como límites. Mucho del análisis superior consiste de una sucesión de ejemplos de este hecho que se hará más claro progresivamente en los siguientes capítulos. La representación de los números irracionales como límites de números racionales puede ser interpretada como el primero y típico ejemplo.

Toda sucesión *convergente* de números conocidos $a_1, a_2, \ldots$ *define* un número l, su límite. Sin embargo, la única prueba de convergencia que surge de la definición de convergencia consiste en estimar las diferencias $|a_n - l|$, y esto es aplicable solamente si el número l es ya conocido. Es esencial disponer de criterios *"intrínsecos"* para la convergencia que no requieran de un conocimiento *a priori* del valor del límite sino que involucren solamente los términos de la propia sucesión. La más simple de tales pruebas o criterios se aplica a una clase especial de sucesiones, las sucesiones monótonas, e incluye muchos de los ejemplos importantes.

Límites de sucesiones monótonas

Una sucesión $a_1, a_2, \ldots$ se denomina *monótonamente creciente* si cada término a_n es mayor o no menor que el precedente; esto es,

$$a_n \geq a_{n-1}.$$

Análogamente, la sucesión es *monótonamente decreciente* si $a_n \leq a_{n-1}$ para todo n. Una sucesión *monótona* es aquélla que es monótonamente creciente o bien monótonamente decreciente. Con esta definición se tiene el principio básico:

Una sucesión que es tanto monótona como acotada converge.[1]

Este principio es sugerido convincentemente, mas no probado, por la intuición; está íntimamente relacionado con las propiedades de los números reales, y es de hecho equivalente al axioma de continuidad para los números reales.

El axioma (ver sección $1b$) de que toda sucesión encaje de intervalos contiene un punto, se ve fácilmente como consecuencia de la convergencia de sucesiones acotadas monótonas. Así, sea $[a_1, b_1], [a_2, b_2], \ldots$ una sucesión de intervalos en encaje. Por la definición de encaje se tiene

$$a_1 \leq a_2 \leq \cdots \leq a_n < b_n \leq b_{n-1} \leq \cdots \leq b_1.$$

Obviamente, la sucesión $a_1, a_2, \ldots$ es monótonamente creciente. También es acotada puesto que $a_1 \leq a_n \leq b_1$ para todo n. De aquí que $l = \lim_{n\to\infty} a_n$ exista. Además, para todo m y para todo número $n > m$ se tiene

$$a_m \leq a_n \leq b_m.$$

Por lo tanto también

$$a_m \leq \lim_{n\to\infty} a_n = l \leq b_m.$$

De aquí que todos los intervalos del encaje anidado contengan uno y el mismo punto l. (El que no tengan otro punto en común se sigue de la propiedad adicional $\lim (b_n - a_n) = 0$ de los encajes de intervalos.)

Criterio de Cauchy para la convergencia

Una sucesión convergente es automáticamente acotada, pero no tiene por qué ser monótona (ver ejemplo b, p. 85). Por lo tanto, al tratar con sucesiones generales es deseable disponer de una prueba para la conver-

[1] La hipótesis sobre la acotación es esencial puesto que ninguna sucesión no acotada puede converger. Obsérvese que una sucesión monótona creciente $a_1, a_2, \ldots$ está siempre "acotada desde abajo": $a_n \geq a_1$ para todo n. A fin de probar que una sucesión monótona creciente converge, es suficiente entonces encontrar un número M tal que $a_n \leq M$ para todo n.

gencia que sea también aplicable a sucesiones no monótonas. Esta necesidad se satisface mediante una simple condición, la *prueba o criterio de convergencia de Cauchy*. Este criterio caracteriza a las sucesiones de números reales que tienen un límite; y lo más importante es que no requiere un conocimiento *a priori* del valor del límite: *Una condición necesaria y suficiente para la convergencia de una sucesión* $a_1, a_2, \ldots$ *es que los elementos* a_n *de la sucesión con un índice n suficientemente grande difieran arbitrariamente poco unos de otros*. Formulado de manera precisa: Una sucesión a_n es convergente si para todo $\varepsilon > 0$ existe un número natural $N = N(\varepsilon)$ tal que $|a_n - a_m| < \varepsilon$ siempre que $n > N$ y $m > N$. Geométricamente, la condición de Cauchy establece que una sucesión converge si existen intervalos arbitrariamente pequeños fuera de los cuales se encuentran solamente un número finito de puntos de la sucesión. La exactitud del criterio de Cauchy para la convergencia será demostrada y discutido su significado en el suplemento.

d. Series infinitas y el símbolo de sumatoria

Una sucesión es simplemente un arreglo ordenado infinito de números $a_1, a_2, \ldots$ *Una serie infinita*

$$a_1 + a_2 + a_3 + \cdots$$

requiere que los términos sean sumados en el orden en el cual aparecen. Para llegar a un significado preciso de la *suma* de una serie infinita se considera la *n-ésima suma parcial,* esto es, la suma de los primeros n términos de la serie

$$s_n = a_1 + a_2 + \cdots + a_n.$$

Las sumas parciales s_n para n diferentes forman una sucesión

$$s_1 = a_1, \qquad s_2 = a_1 + a_2, \qquad s_3 = a_1 + a_2 + a_3,$$

y así sucesivamente. La suma s de la serie infinita está definida entonces como

$$s = \lim_{n\to\infty} s_n,$$

siempre que el límite exista. En ese caso la serie infinita se denomina *convergente*. Si la sucesión s_n diverge, la serie infinita se denomina *divergente*. Por ejemplo, la sucesión $1, q, q^2, q^3, \ldots$ da lugar a la serie geométrica infinita

$$1 + q + q^2 + q^3 + \cdots$$

cuyas sumas parciales son

$$s_n = 1 + q + q^2 + \cdots + q^{n-1}.$$

Para $|q| < 1$ la sucesión s_n converge hacia el límite

$$s = \frac{1}{1 - q},$$

el cual representa entonces la suma de la serie infinita. Para $|q| \geq 1$ las sumas parciales s_n no poseen límite y la serie diverge (ver pp. 90-91).

Es costumbre utilizar para $a_1 + a_2 + \cdots + a_n$ el símbolo

$$\sum_{k=1}^{n} a_k$$

que indica que la suma de los a_k ha de tomarse con k recorriendo los enteros $k = 1$ a $k = n$. Por ejemplo,

$$\sum_{k=1}^{4} \frac{1}{k!} \text{ es el símbolo para } \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!},$$

mientras que

$$\sum_{k=1}^{n} a^k b^{2k} \text{ es el símbolo para } a^1 b^2 + a^2 b^4 + a^3 b^6 + \cdots + a^n b^{2n}.$$

Más generalmente, $\sum_{k=m}^{n} a_k$ significa la suma de todo a_k obtenido dando a k los valores m, $m + 1$, $m + 2$, ... , n. Por tanto

$$\sum_{k=3}^{5} \frac{1}{k!} = \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!}.$$

En estos ejemplos se ha utilizado la letra k para el índice de sumatoria. Por supuesto, la suma es independiente de la letra que denota este índice. Por lo tanto

$$s_n = \sum_{k=1}^{n} a_k = \sum_{i=1}^{n} a_i.$$

El símbolo

$$\sum_{k=1}^{\infty} a_k$$

se utiliza para denotar la suma de toda la serie infinita. Análogamente, $\sum_{k=0}^{\infty} a_k$ sería el símbolo para la suma de la serie infinita $a_0 + a_1 + a_2 + \ldots$, cuya n-ésima suma parcial es $s_n = a_0 + a_1 + a_2 + \cdots + a_{n-1}$.

Muchos de los resultados anteriores pueden ser escritos más concisamente en esta notación sumatoria. La fórmula de la p. 81 para la suma de los primeros n cuadrados se convierte en

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$$

La fórmula para la suma de la serie geométrica es

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q} \qquad \text{para } |q| < 1.$$

Finalmente, el teorema binomial está expresado por

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k.$$

Puesto que una serie infinita es simplemente el límite de una sucesión s_n, la convergencia puede ser decidida sobre la base de los criterios de convergencia para sucesiones. Por ejemplo, la convergencia de la serie

$$\sum_{k=1}^{\infty} \frac{1}{k^k} = \frac{1}{1^1} + \frac{1}{2^2} + \frac{1}{3^3} + \cdots$$

se sigue inmediatamente del hecho de que las sumas parciales

$$s_n = \sum_{k=1}^{n} \frac{1}{k^k} = \frac{1}{1^1} + \frac{1}{2^2} + \frac{1}{3^3} + \cdots + \frac{1}{n^n}$$

crecen monótonamente con n y están acotadas puesto que

$$1 \leq s_n \leq 1 + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \cdots + \frac{1}{2^n}$$

$$= 1 + \frac{1}{4}\frac{1 - 1/2^{n-1}}{1 - \frac{1}{2}} = 1 + \frac{1}{2} - \frac{1}{2^n} < \frac{3}{2}.$$

Posteriormente, en el capítulo 7, se estudiarán las series infinitas más sistemáticamente.

e. El número *e*

Como primer ejemplo de un número que es generado como el límite de una sucesión, se considera

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots.$$

Así e es la notación para $\lim_{n\to\infty} S_n$, donde

$$S_n = 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!}$$ [1]

[1] Recordando la convención que define a 0! como 1, el primer término de la serie puede escribirse como 1/0! en concordancia con la ley de formación de los siguientes términos. Nótese que en nuestra notación S_n es realmente la $(n+1)$-ésima suma parcial de la serie infinita, en vez de la n-ésima. Esto, sin embargo, no tiene ningún significado.

Los números e y π son las constantes trascendentes más exactamente utilizadas en análisis matemático. Para demostrar la existencia del límite e se requiere solamente probar que la sucesión S_n es acotada, puesto que los números S_n crecen monótonamente. Para todos los valores de n se tiene

$$\begin{aligned} S_n &= 1 + 1 + \frac{1}{2} + \frac{1}{2\cdot 3} + \frac{1}{2\cdot 3\cdot 4} + \cdots + \frac{1}{2\cdot 3\cdot 4\cdot \cdots \cdot n} \\ &\leq 1 + 1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots + \frac{1}{2^{n-1}} \\ &= 1 + \frac{1 - 1/2^n}{1 - \frac{1}{2}} < 3. \end{aligned}$$

Los números S_n tienen por lo tanto la cota superior 3, y puesto que forman una sucesión monótona creciente los números poseen un límite, que se denota por e.

La expresión de e como una serie permite rápidamente calcular e con gran precisión. El error que se comete al aproximar e por una suma parcial S_m puede ser estimado por el mismo método de comparación con una serie geométrica que proporcionó la cota superior 3 para e. Para toda $n > m$ se tiene

$$\begin{aligned} S_n &= S_m + \frac{1}{(m+1)!} + \frac{1}{(m+2)!} + \cdots + \frac{1}{n!} \\ &\leq S_m + \frac{1}{(m+1)!}\left[1 + \frac{1}{m+2} + \frac{1}{(m+2)(m+3)} + \cdots\right] \\ &\leq S_m + \frac{1}{(m+1)!}\left[1 + \frac{1}{m+1} + \frac{1}{(m+1)^2} + \cdots\right] \\ &= S_m + \frac{1}{(m+1)!}\frac{1}{1 - \dfrac{1}{m+1}} = S_m + \frac{1}{m}\frac{1}{m!}. \end{aligned}$$

Por tanto, para $n > m$

$$S_m < S_n \leq S_m + \frac{1}{m}\frac{1}{m!}.$$

Si se hace crecer n más allá de toda cota mientras a m se le mantiene fijo, se encuentra también que

$$S_m < e \leq S_m + \frac{1}{m}\frac{1}{m!}.$$

De aquí que e difiere de S_m en a lo más $(1/m)(1/m!)$. Puesto que $m!$ crece muy rápidamente con m, el número S_m es ya una buena aproximación para m bastante pequeño; por ejemplo, S_{10} difiere de e en menos que 10^{-7}. De esta manera se encuentra que $e = 2.718281\cdots$

e es un *número irracional.* La estimación para e en términos de S_m puede utilizarse también para establecer esta propiedad. Así, si e fuera racional, e podría escribirse en la forma p/m con p y m enteros positivos; aquí $m \geq 2$, puesto que e, situado entre 2 y 3, no puede ser un entero. Comparando e con la suma parcial S_m, se tendría

$$S_m < \frac{p}{m} \leq S_m + \frac{1}{m}\frac{1}{m!}.$$

Si se multiplican aquí los dos miembros por $m!$ se encuentra que

$$m!\,S_m < p(m-1)! \leq m!\,S_m + \frac{1}{m} < m!\,S_m + 1.$$

Pero

$$m!\,S_m = m! + m! + \frac{m!}{2!} + \frac{m!}{3!} + \cdots + \frac{m!}{m!}$$

es un entero puesto que cada término de la suma lo es. Por lo tanto, si e fuera racional el entero $p(m-1)$ estaría situado entre dos enteros sucesivos, lo cual es imposible.[1]

El número e como límite de $(1 + 1/n)^n$. El número e, que se definió aquí como la suma de una serie infinita, puede obtenerse también como límite de la sucesión

$$T_n = \left(1 + \frac{1}{n}\right)^n.$$

La demostración es sencilla y es al mismo tiempo un ejemplo instructivo de operaciones con límites. De acuerdo con el teorema del binomio,

$$\begin{aligned} T_n &= \left(1 + \frac{1}{n}\right)^n \\ &= 1 + n\frac{1}{n} + \frac{n(n-1)}{2!}\frac{1}{n^2} + \cdots + \frac{n(n-1)(n-2)\cdots 1}{n!}\frac{1}{n^n} \\ &= 1 + 1 + \frac{1}{2!}\left(1 - \frac{1}{n}\right) + \cdots \\ &\qquad + \frac{1}{n!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{n-1}{n}\right) \end{aligned}$$

[1] La irracionalidad del número e significa que no existe ninguna ecuación lineal $ax + b = 0$ con coeficientes racionales a, b y $a \neq 0$ que tenga a e como solución. Una proposición mucho más fuerte ha sido probada (por Hermite), a saber: que no existe ninguna ecuación polinomial $a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n = 0$, de cualquier grado n y con coeficientes racionales $a_0, a_1, \ldots, a_n$ (con $a_0 \neq 0$), con $x = e$ como raíz. Se dice que e es un número *trascendente,* en contraste con los números "algebraicos" como $\sqrt{2}$ o bien $\sqrt[3]{10}$, que son raíces de ciertas ecuaciones polinomiales con coeficientes racionales.

De esto se ve de inmediato que $T_n \leq S_n < 3$. Más aún, puesto que T_{n+1} se obtiene de T_n reemplazando los factores $1 - 1/n$, $1 - 2/n$, ... por los factores mayores $1 - 1/(n+1)$, $1 - 2/(n+1)$, ..., y finalmente añadiendo un término positivo, se ve que los T_n forman también una sucesión monótona creciente de la cual se sigue la existencia del límite $\lim_{n\to\infty} T_n = T$. Para probar que $T = e$, se observa que para $m > n$

$$T_m > 1 + 1 + \frac{1}{2!}\left(1 - \frac{1}{m}\right) + \cdots + \frac{1}{n!}\left(1 - \frac{1}{m}\right)\cdots\left(1 - \frac{n-1}{m}\right).$$

Si ahora se mantiene n fijo y se deja crecer m más allá de toda cota, se obtiene en el miembro izquierdo el número T y en el derecho la expresión S_n, de modo que $T \geq S_n$. Por lo tanto $T \geq S_n \geq T_n$ para todo valor de n. Se hace ahora crecer n, de modo que T_n tiende a T; y de la doble desigualdad se sigue que $T = \lim_{n\to\infty} S_n = e$. Esta era la propiedad por demostrar.

Posteriormente (sección 2.6, p. 171) este número e se obtendrá aún nuevamente desde otro punto de vista.

f. El número π como límite

Un proceso límite que en esencia se remonta a la antigüedad clásica (Arquímedes) es aquél mediante el cual se define el número π. Geométricamente π significa el área de un círculo de radio uno. Se considera como cosa obvia que esta área puede ser expresàda por un número (racional o irracional) denotado por π. Sin embargo, esta definición no es de gran ayuda para nosotros si se desea calcular el número con toda precisión. Entonces no se tiene más alternativa que la de representar el número por medio de un proceso límite, a saber, como el límite de una sucesión de números conocidos y fácilmente calculados. Arquímedes utilizó ya este proceso en su método de exhaucción, el cual consiste en aproximar el círculo por medio de polígonos regulares con un número creciente de lados que se ajustan a él cada vez más. Si f_m denota el área de un m-ágono (polígono de m lados) regular inscrito en el círculo, el área del $2m$-ágono inscrito está dado por la fórmula [probada por geometría elemental o bien por la expresión $f_n = (n/2)$ sen $(2\pi/n)$ (ver Fig. 1.46)]

$$f_{2m} = \frac{m}{2}\sqrt{2 - 2\sqrt{1 - (2f_m/m)^2}}.$$

Déjese recorrer ahora a m, no la sucesión de todos los enteros positivos, sino por la sucesión de potencias de 2, esto es, $m = 2^n$; en otras palabras, se forman aquellos polígonos regulares cuyos vértices se obtienen mediante

bisección repetida de la circunferencia. Es claro de la interpretación geométrica que los f_{2^n} forman una sucesión creciente y acotada, que por lo tanto *tiene* un límite que es el área del círculo:

$$\pi = \lim_{n\to\infty} f_{2^n}.$$

Esta representación de π como límite sirve en realidad como base para cálculos numéricos; así, partiendo con el valor $f_4 = 2$ pueden calcularse en orden los términos de la sucesión que tienden a π. Una estimación de la precisión con la cual cualquier término f_{2^n} representa a π puede obtenerse construyendo las rectas que tocan al círculo y son paralelas a los lados del 2^n-ágono inscrito. Estas rectas forman un polígono circunscrito, análogo al 2^n-ágono inscrito, que tiene dimensiones mayores en la razón $1 : \cos(\pi/2^n)$. Por lo tanto, el área F_{2^n} del polígono circunscrito puede encontrarse a partir de la razón dada por

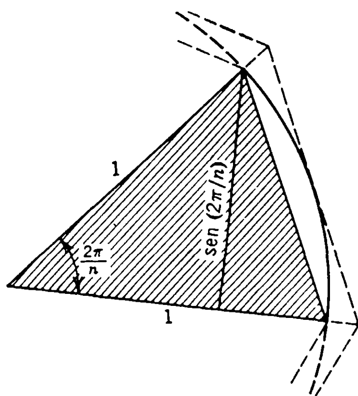


Figura 1.46.

$$\frac{f_{2^n}}{F_{2^n}} = \left(\cos\frac{\pi}{2^n}\right)^2.$$

Puesto que el área del polígono circunscrito es mayor que la del círculo, se tiene

$$f_{2^n} < \pi < F_{2^n} = \frac{f_{2^n}}{\left(\cos\frac{\pi}{2^n}\right)^2} = \frac{2f_{2^n}}{1+\sqrt{1-(f_{2^n}/2^{n-1})^2}}.$$

Por ejemplo, $f_8 = 2\sqrt{2}$, de modo que se tiene la estimación

$$2\sqrt{2} < \pi < \frac{4\sqrt{2}}{1+\frac{1}{2}\sqrt{2}}.$$

Estas son cuestiones con las cuales el lector estará más o menos familiarizado. Sin embargo, lo que se desea hacer resaltar es que el cálculo de áreas por medio de exahucción por figuras rectilíneas cuyas áreas pueden ser calculadas fácilmente constituye la base para el concepto de integral, el que se introducirá en el capítulo 2. Para el cálculo numérico efectivo de π se dispone de métodos mucho más eficientes como se verá en la sección 6.26.

1.8 El concepto de límite para funciones de una variable continua

Hasta ahora se han considerado límites de sucesiones, esto es, de funciones de una variable entera n. La noción de límite, sin embargo, se encuentra frecuentemente en relación con una función $f(x)$ que está definida para toda x en algún intervalo.

Se dice que el valor de la función $f(x)$ tiende a un límite η cuando x tiende a ξ, o sea, en símbolos,

$$\lim_{x \to \xi} f(x) = \eta,$$

si $f(x)$ difiere arbitrariamente poco de η para toda x para la cual $f(x)$ está definida y que está situada suficientemente cerca de ξ.[1] Expresada más precisamente, la definición de $\lim f(x)$ es como sigue.

Si para una cantidad positiva arbitraria ε asignada puede definirse un intervalo $|x - \xi| < \delta$ suficientemente pequeño de modo que para toda x que pertenece tanto al dominio de f como al intervalo la desigualdad $|f(x) - \eta| < \varepsilon$ se satisface, entonces $\lim_{x \to \xi} f(x) = \eta$.

Existe una conexión muy estrecha entre los conceptos de límite de una función y continuidad. Si ξ pertenece al dominio de f, esto es, si $f(\xi)$ está definido, entonces $\lim_{x \to \xi} f(x)$, si existe, debe tener el valor $f(\xi)$. Realmente, la definición de $\eta = \lim_{x \to \xi} f(x)$ implica en particular $|f(\xi) - \eta| < \varepsilon$ para todo ε positivo, y por lo tanto $\eta = f(\xi)$. Ahora, comparando las definiciones de límite y de continuidad se ve que la relación $\lim_{x \to \xi} f(x) = f(\xi)$ justamente expresa la continuidad de la función f en el punto ξ. Luego, para ξ en el dominio de f, la existencia de $\lim_{x \to \xi} f(x)$ significa justamente que f es continua en ξ. Más generalmente, si $f(x)$ no está definida en ξ pero $\lim_{x \to \xi} f(x)$ existe y tiene el valor η, puede asignarse a f en el punto ξ el valor η, y la función f, completada de este modo, será continua en ξ. (Singularidad Removible. Ver p. 59.)

El *límite de una función* puede ser también completamente descrito en términos de *límites de sucesiones.* El aserto

$$\lim_{x \to \xi} f(x) = \eta$$

significa que

$$\lim_{n \to \infty} f(x_n) = \eta$$

para toda sucesión x_n con límite ξ (donde, claro está, se supone que los x_n pertenecen al dominio de f). Pues si $\lim_{x \to \xi} f(x) = \eta$ y si $\lim_{n \to \infty} x_n = \xi$, entonces $f(x)$ está

---

[1] Se supone aquí que arbitrariamente cerca de ξ existen puntos donde f está definida.

arbitrariamente cerca de η para x suficientemente cercano a ξ; pero x_n está suficientemente cercano a ξ sólo si n es suficientemente grande, y consecuentemente $\lim_{n\to\infty} f(x_n) = \eta$. Si, por otra parte, $\lim_{n\to\infty} f(x_n) = \eta$ para $x_n \to \xi$, debe tenerse también $\lim_{x\to\xi} f(x) = \eta$. Pues de lo contrario existiría un número positivo ε tal que $|f(x) - \eta| \geq \varepsilon$ para algún x arbitrariamente cercano a ξ; también existiría entonces una sucesión x_n convergente a ξ para la cual $|f(x_n) - \eta| \geq \varepsilon$; pero entonces $\lim_{n\to\infty} f(x_n)$ no podría ser igual a η.

La continuidad de la función $f(x)$ en el punto ξ implica entonces: $\lim_{n\to\infty} f(x_n) = f(\xi)$, para toda sucesión x_n en el dominio de f que converge a ξ. Más generalmente, para una función continua en el *intervalo* la relación

$$\lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n)$$

es válida para cualquier sucesión en el dominio de f que converge a un punto del intervalo. Se ve que para una función continua el símbolo de límite puede ser intercambiado (o bien, como se dice, "conmuta") con los símbolos para la función.

Límites de sumas, productos y cocientes de funciones se determinan por las mismas reglas que para las sucesiones (ver p. 94): Si $\lim_{x\to\xi} f(x) = \eta$ y $\lim_{x\to\xi} g(x) = \zeta$ existen, entonces

$$\lim_{x\to\xi} (f(x) + g(x)) = \eta + \zeta, \qquad \lim_{x\to\xi} (f(x)g(x)) = \eta\zeta;$$

y para $\zeta \neq 0$ también

$$\lim_{x\to\xi} \frac{f(x)}{g(x)} = \frac{\eta}{\zeta}$$

Las demostraciones son las mismas que para sucesiones. (Las reglas se seguirían también de aquéllas para sucesiones, escribiendo límites de funciones como límites de sucesiones.) Consecuentemente, cuando ξ pertenece al dominio de f y g, *la suma, producto y cociente de dos funciones $f(x)$ y $g(x)$ que son continuas en un punto ξ son también continuas* (donde para los cocientes se debe suponer que $g(\xi) \neq 0$).

Los casos en que ξ no pertenece al dominio de f resultarán ser de particular importancia para el cálculo diferencial. Como un primer ejemplo, considérese la relación

$$\lim_{x\to\xi} \frac{x^n - \xi^n}{x - \xi} = n\xi^{n-1},$$

donde n es un entero positivo. Por supuesto, $f(x) = (x^n - \xi^n)/(x - \xi)$ es una función definida solamente para $x \neq \xi$. Pero para $x \neq \xi$ la identidad algebraica

$$\frac{x^n - \xi^n}{x - \xi} = x^{n-1} + x^{n-2}\xi + x^{n-3}\xi^2 + \cdots + \xi^{n-1},$$

es válida como consecuencia de la fórmula sumatoria para la serie geométrica. Para encontrar el límite se requiere solamente hacer tender x a ξ y evaluar el límite del lado derecho por las reglas para límites de sumas y cocientes.

Menos obvia es la fórmula

$$\lim_{x\to 0} \frac{\operatorname{sen} x}{x} = 1$$

(donde, por supuesto, el ángulo x es medido en "radianes", como se explicó en las pp. 73-74). Aquí también, el cociente $(\operatorname{sen} x)/x$ está definido solamente para $x \neq 0$.

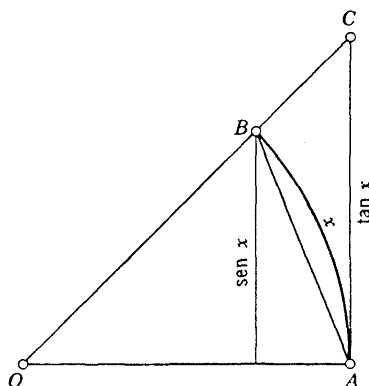


Figura 1.47.

Pero si se *define* $(\operatorname{sen} x)/x = 1$ para $x = 0$, el cociente se completa como una función que es continua también en $x = 0$. Para la demostración de la fórmula límite se apela aquí a un argumento geométrico.

De la Fig. 1.47 se encuentra, comparando las áreas de los triángulos OAB y OAC y el sector OAB[1] del círculo unitario, que, si $0 < x < \pi/2$, es

$$\tfrac{1}{2} \operatorname{sen} x < \tfrac{1}{2} x < \tfrac{1}{2} \tan x.$$

De aquí se sigue que si $0 < |x| < \pi/2$, es

$$1 < \frac{x}{\operatorname{sen} x} < \frac{1}{\cos x}.$$

Por lo tanto, el cociente $(\operatorname{sen} x)/x$ está situado entre los números 1 y $\cos x$. Se sabe que $\cos x$ tiende a 1 cuando $x \to 0$, y de esto se sigue que el cociente $(\operatorname{sen} x)/x$ puede sólo diferir arbitrariamente poco de 1 si x está

[1] Por supuesto, el ángulo x pudo haberse definido en primer término como el doble del área del sector OAB.

suficientemente cerca de 0. Esto es exactamente lo que se entiende por la ecuación que había que demostrar.

Del resultado acabado de probar se sigue que

$$\lim_{x\to 0}\frac{\tan x}{x} = \lim_{x\to 0}\frac{\operatorname{sen} x}{x}\lim_{x\to 0}\frac{1}{\cos x} = 1,$$

y también

$$\lim_{x\to 0}\frac{1-\cos x}{x} = 0.$$

Esto último se sigue de la fórmula, válida para $0 < |x| < \pi/2$,

$$\frac{1-\cos x}{x} = \frac{(1-\cos x)(1+\cos x)}{x(1+\cos x)} = \frac{1-\cos^2 x}{x(1+\cos x)}$$

$$= \frac{\operatorname{sen} x}{x}\cdot\frac{1}{1+\cos x}\cdot \operatorname{sen} x.$$

Para $x \to 0$ el primer factor en el lado derecho tiende a 1, el segundo a $\frac{1}{2}$, y el tercero a 0, como se estableció antes.

Dividiendo la misma fórmula por x, se obtiene

$$\frac{1-\cos x}{x^2} = \left(\frac{\operatorname{sen} x}{x}\right)^2\frac{1}{1+\cos x},$$

de la cual surge

$$\lim_{x\to 0}\frac{1-\cos x}{x^2} = \frac{1}{2}$$

Límites para $x \to \infty$. Finalmente se hace notar que igualmente es posible considerar procesos límite en los cuales la variable continua x crece más allá de toda cota. Por ejemplo, es claro el significado de la ecuación

$$\lim_{x\to\infty}\frac{x^2+1}{x^2-1} = \lim_{x\to\infty}\frac{1+1/x^2}{1-1/x^2} = 1.$$

Significa que la función en la izquierda difiere arbitrariamente poco de uno, bastando sólo con que x sea suficientemente grande. Las reglas para formar límites de este tipo para sumas, productos y cocientes son las mismas que antes.

* Existe un resultado adicional el cual es frecuentemente útil en el cálculo de límites, y es la regla para obtener el límite de una función compuesta. La función *compuesta* $f(g(z))$ está definida para aquellos valores de z para los cuales $x = g(z)$ está situado en el dominio de $f(x)$. La función $g(z)$ puede ser una función de una variable continua o bien una variable entera, pero $f(x)$ debe ser una función de una variable continua.

Si $\lim_{z \to \zeta} g(z) = \xi$, donde ξ está situado dentro de un intervalo abierto del dominio de f, y si $\lim_{x \to \xi} f(x) = \eta$, entonces $\lim_{z \to \zeta} f(g(z)) = \eta$. Como un corolario, se observa que una función continua de una función continua es asimismo continua (como se mencionó ya en la p. 78).

El resultado es obvio del hecho de que $f(x)$ puede hacerse arbitrariamente cercano a η tomando x suficientemente cercano a ξ, y para hacer $x = g(z)$ suficientemente cercano a ξ bastará solamente con tomar z suficientemente cercano a ζ. Con algunas ligeras modificaciones se aplican los mismos argumentos que valen en el caso en que cualquiera de las variables puede crecer más allá de toda cota.

a. Algunas observaciones sobre las funciones elementales

Hasta aquí tácitamente se ha supuesto que las funciones elementales son continuas. La demostración de este hecho es muy simple. En primer lugar, la función $f(x) = x$ es continnua; por ello, $x^2 = x \cdot x$ es continua, como producto de dos funciones continuas, y toda potencia de x es, de la misma manera, continua. Por lo tanto, todo polinomio es continuo, por ser la suma de funciones continuas. Toda función racional, como cociente de funciones continuas, es análogamente continua en todo intervalo en el cual el denominador no se hace cero.

La función x^n es continua y monótona para $x \geq 0$. Por lo tanto, la n-ésima raíz, que es la función inversa de la n-ésima potencia, es continua. De este hecho es fácil concluir que la n-ésima raíz de una función racional es continua (excepto donde el denominador se hace cero).

La continuidad de las funciones trigonométricas puede ser probada ahora, utilizando los conceptos ya desarrollados. Sin embargo, la discusión se omite aquí puesto que en el capítulo 2 (p. 188) se verá que la continuidad de todas estas funciones se sigue fácilmente como consecuencia de su diferenciabilidad.

Se harán solamente algunos comentarios en torno a la definición y continuidad de la función exponencial a^x, de la función potencia general, x^α, y del logaritmo. Se supone, como en la sección 1.3 (p. 74), que a es un número positivo, digamos, mayor que uno, y $r = p/q$ es un número racional positivo (siendo p y q enteros); entonces $a^r = a^{p/q}$ es el número positivo cuya q-ésima potencia es a^p. Si α es cualquier número irracional y $r_1, r_2, \ldots, r_m, \ldots$ es una sucesión de números racionales que se aproximan a α, se afirma que $\lim_{m \to \infty} a^{r_m}$ existe; y este límite se denomina entonces a^α.

Para probar la existencia de este límite por medio del criterio de Cauchy se requiere mostrar solamente que $|a^{r_n} - a^{r_m}|$ es arbitrariamente pequeño siempre que n y m sean suficientemente grandes. Se supone, por ejemplo, que $r_n > r_m$, o bien que $r_n - r_m = \delta$, donde $\delta > 0$. Entonces,

$$a^{r_n} - a^{r_m} = a^{r_m}(a^\delta - 1).$$

Puesto que los r_m convergen a α, éstos son acotados y también lo son los a^{r_m}; por lo tanto, es suficiente mostrar que

$$|a^\delta - 1| = a^\delta - 1$$

es arbitrariamente pequeño cuando los valores de n y m son suficientemente grandes. Sin embargo, ciertamente el número racional δ puede hacerse tan pequeño como se desee siempre que los valores de n y m sean suficientemente grandes. Por lo tanto, si l es un entero positivo arbitrariamente grande, $\delta < 1/l$ si n y m son suficientemente grandes. Ahora las relaciones $\delta < 1/l$ y $a > 1$ dan [1]

$$1 < a^\delta < a^{1/l},$$

y puesto que $a^{1/l}$ tiende a uno conforme l crece hacia infinito (ver p. 87), se concluye inmediatamente que vale el aserto y el límite existe.

Puede mostrarse que la función a^x extendida de esta manera a valores irracionales es también continua en todas partes, y, aún más, que es monótona. Para valores negativos de x esta función está naturalmente definida por la ecuación

$$a^x = \frac{1}{a^{-x}}.$$

Conforme x varía de $-\infty$ a $+\infty$, a^x toma valores entre cero y $+\infty$. Consecuentemente, posee una función inversa que es continua y monótona, la cual se denomina el *logaritmo de base a.* De manera parecida puede probarse que la potencia general x^α es una función continua de x, donde α es cualquier número fijo racional o irracional y x varía en el intervalo $0 < x < \infty$, la cual es monótona si $\alpha \neq 0$.

La discusión "elemental" de la función exponencial, del logaritmo y de la potencia x^α, delineada aquí será reemplazada posteriormente (p. 171) por otra discusión en principio mucho más simple.

Suplemento

Uno de los grandes logros de los matemáticos griegos fue la reducción de asertos matemáticos y teoremas, en forma lógica y coherente, a un pequeño número de postulados o axiomas muy simples, los bien conocidos axiomas de la geometría, o bien las reglas de la aritmética, que gobiernan relaciones entre unos cuantos objetos básicos, tales como los números enteros y los puntos geométricos. Los objetos matemáticos básicos se originaron como abstracciones o idealizaciones de la realidad física. Los axiomas, ya sea considerados como "evidentes" desde un punto de vista filosófico o bien meramente como abrumadoramente plausibles, se aceptan sin demostración; sobre ellos reposa la cristalizada estructura de las matemáticas. Por muchos siglos la axiomática matemática euclidiana fue aceptada como un modelo para el estilo matemático, y aun imitada para otras tareas intelectuales. (Por ejemplo, filósofos tales como Descartes y Spinoza trataron de hacer sus especulaciones más convincentes presentándolas axiomáticamente, o bien, como ellos decían, "more geometrico".)

---

[1] Esta propiedad surge del hecho de que para $a > 1$ la potencia $a^{m/n}$ es mayor que 1 si m/n es positivo. Pues $a = (a^{1/n})^m$ es el producto de m factores todos mayores que 1 y así es mayor que uno.

El método axiomático fue descartado cuando, después del estancamiento durante el medioevo, las matemáticas junto con la ciencia natural principiaron un desarrollo explosivamente vigoroso basado en el nuevo cálculo. Extendiendo vastamente el alcance de las matemáticas, los descubridores ingeniosos no podían ser estorbados teniendo que sujetar los nuevos descubrimientos a un análisis lógico riguroso, y por ello en el siglo XVII la invocación a la evidencia intuitiva se convirtió en un substituto ampliamente utilizado en cambio de la demostración deductiva. Matemáticos de primer orden operaron con los nuevos conceptos guiados por un infalible sentido para la exactitud de los resultados, aun, en ocasiones, con místicas asociaciones, como en las referencias a los "infinitesimales" o "cantidades infinitamente pequeñas". La confianza en la arrolladora potencia de las nuevas manipulaciones del cálculo transportó a los investigadores muy lejos, en trayectorias imposibles de seguir si hubiesen estado sujetos a las limitaciones del complejo rigor lógico. Solamente el instinto seguro de grandes maestros pudo prevenir en contra de graves errores.

El entusiasmo poco crítico pero enormemente fructífero del primer período se fue encontrando gradualmente con contracorrientes que se elevaron con gran fuerza en el siglo XIX pero que no impidieron el desarrollo del análisis constructivo iniciado anteriormente. Muchos de los grandes matemáticos del siglo XIX, en particular Cauchy y Weierstrass, desempeñaron un papel en el esfuerzo hacia la crítica revaluación. El resultado no fue sólo un nuevo y firme fundamento del análisis, sino que también crecieron la lucidez y la simplicidad como bases para notables progresos posteriores.

Una meta importante era la de reemplazar la confianza sin distinción sobre la "intuición" imprecisa por un razonamiento preciso basado en operaciones con números, pues el pensamiento geométrico intuitivo deja un margen indeseable de vaguedad, como se verá con frecuencia en los siguientes capítulos. Por ejemplo, el concepto general de curva continua escapa a la intuición geométrica. Una curva continua, que representa una función continua, como se definió anteriormente, puede no tener una dirección definida en cada punto; más aún, pueden construirse funciones continuas cuyas gráficas no tienen dirección en ninguna parte o bien a las cuales no puede asignárseles ninguna longitud.

Empero, jamás debe olvidarse que el razonamiento abstracto deductivo es meramente un aspecto de las matemáticas, mientras que la motivación impulsora y el gran alcance universal del análisis emanan de la realidad física y de la geometría intuitiva.

Este suplemento proporcionará un apoyo riguroso (con algunas repeticiones) para los conceptos básicos tratados intuitivamente antes en este capítulo.

S.1 Los límites y el concepto de número

Empezaremos con las ideas de la sección 1.1, analizando de lleno el concepto de número real y su conexión con el de límite. Se define el continuo de números por un procedimiento constructivo basado en los números naturales. Se probará entonces que el concepto extendido de número satisface las reglas de aritmética y los otros requisitos, haciendo de éste la herramienta adecuada para la medición.

Puesto que una exposición completa requeriría un libro separado,[1] se indicarán solamente los pasos principales. El estudiante, esforzándose a través del material algo tedioso, se maravillará del hecho de que sobre la base de los números naturales la mente humana puede erigir un sistema de números lógico y consistente, formidablemente adecuado para la labor de medición científica.[2]

a. Los números racionales

Límites definidos por intervalos racionales. Se comienza aceptando el sistema de números racionales con todas sus propiedades usuales, obtenidas de las propiedades básicas de números naturales. Así, los números racionales están *ordenados* por magnitud, permitiéndose definir intervalos "racionales" como conjuntos de números racionales situados entre dos números racionales dados (los intervalos que incluyen los puntos extremos se denominan *cerrados*). La longitud del intervalo con puntos extremos a, b es $|b - a|$. Como se observa en la sección 1*a*, los números racionales son *densos* y todo intervalo racional contiene una infinidad de números racionales. Por el momento todas las cantidades que ocurran se supondrán números racionales.

Dentro del dominio de los números racionales se definen sucesiones y límites (ver p. 93). Dada una sucesión infinita de números racionales $a_1, a_2, \ldots$ y un número *racional* r se dice que

$$\lim_{n\to\infty} a_n = r$$

[1] Véase por ejemplo, E. Landau, *Foundations of Analysis*, 2nd. Ed., Chelsea, New York, 1960.

[2] Los números reales pueden ser introducidos también de manera puramente axiomática, con todas sus propiedades básicas aceptadas como *axiomas*. En el tratamiento que se hará aquí se aceptan, en principio, solamente los axiomas para números naturales (incluyendo el principio de inducción matemática). Los números racionales y los números reales son *construidos* entonces sobre esta base. Los "axiomas" para números reales son entonces, en principio, simplemente *teoremas* sobre los números naturales, para los cuales se requieren demostraciones. En realidad, se comenzará ya con los números racionales como elementos conocidos, puesto que la construcción de los números racionales a partir de los números naturales y la deducción de las propiedades básicas de los números racionales no presentan ninguna dificultad.

si todo intervalo racional que contiene a r en su interior contiene también a *"casi todo"* a_n, esto es, todos los a_n excepto a lo más un número *finito*. Se sigue de inmediato que una sucesión de números racionales no puede tener más de un límite racional y que las reglas usuales para límites de sumas, diferencias, productos y cocientes (ver pp. 94-95) son válidas para sucesiones de números racionales con límites racionales.

Una consecuencia enteramente obvia de esta definición es que pasando al límite se preserva el orden: si $\lim a_n = a$, $\lim b_n = b$ y para todo n es $a_n \leq b_n$, entonces $a \leq b$. Nótese que aún suponiendo $a_n < b_n$ estrictamente, no puede decirse más que $a \leq b$, o bien excluir la posible igualdad de los límites (por ejemplo, las sucesiones $a_n = 1 - 2/n$ y $b_n = 1 - 1/n > a_n$ poseen el mismo límite 1).

Las aseveraciones en torno a límites pueden ser expresadas en términos de *sucesiones nulas* racionales, esto es, sucesiones $a_1, a_2, \ldots$ de números racionales para los cuales

$$\lim_{n\to\infty} a_n = 0.$$

Se dice a_n "se torna arbitrariamente pequeño conforme n tiende a infinito", dándose a entender con ello que para todo racional positivo ε, no importa cuán pequeño, se satisface la desigualdad $|a_n| < \varepsilon$ para casi todo n. Obviamente la sucesión $a_n = 1/n$ es una sucesión nula.

Así, una sucesión de números racionales a_n tiene el límite racional r si y sólo si los números $r - a_n$ forman una sucesión nula.

b. Números reales determinados por encaje de intervalos racionales

Se observó en la p. 29 que intuitivamente los puntos racionales son densos en el eje real y que existen siempre números racionales entre dos números reales cualesquiera. Esto sugiere la posibilidad de *definir* rigurosamente un número real totalmente en términos de relaciones de orden respecto a los racionales, procedimiento que se seguirá a continuación.

Una "sucesión anidada" o *encaje* de intervalos racionales (pp. 32-33) es una sucesión de intervalos *cerrados* J_n con puntos extremos *racionales* a_n, b_n, con cada intervalo contenido en el precedente, cuyas longitudes forman una sucesión nula

$$a_{n-1} \leq a_n \leq b_n \leq b_{n-1}$$

y

$$\lim_{n\to\infty} (b_n - a_n) = 0.$$

Puesto que cada intervalo $J_n = [a_n, b_n]$ de un encaje contiene todos los intervalos subsiguientes, un número racional r situado fuera de cualquier J_n está situado también fuera y sobre el mismo lado de todos los intervalos

subsiguientes. Así, un encaje de intervalos racionales da lugar a una separación de todos los números racionales en tres clases.[1] La primera clase consiste de los números racionales r situados a la izquierda de los intervalos J_n para n suficientemente grande, o sea, para los cuales $r < a_n$ para casi todo n. La segunda clase consiste de los números racionales r contenidos en todos los intervalos J_n. Esta clase contiene a lo más un número, puesto que la longitud del intervalo J_n se reduce a cero con n creciente. La tercera clase consiste de los números racionales r para los cuales $r > b_n$ para casi todo n. Es claro que todo número de la primera clase es menor que cualquiera de la segunda clase, y cualquier número de la segunda clase es menor que cualquiera de la tercera. Los propios puntos a_n son de la primera o bien de la segunda clase, y los números b_n están en la segunda o bien en la tercera.

Si la segunda clase no es vacía, consiste de un solo número racional r. En este caso la primera clase consiste de los números racionales menores que r y la tercera clase de los números racionales mayores que r. Se dice entonces que el encaje de intervalos J_n *representa* el número racional r. Por ejemplo, el encaje de intervalos $[r - 1/n, r + 1/n]$ representa el número r.

Si la segunda clase es vacía, el encaje no representa un número racional; estas sucesiones anidadas sirven entonces para representar números irracionales. Los intervalos individuales $[a_n, b_n]$ de la sucesión no son importantes para este propósito; solamente la separación de los números racionales en tres clases, generada por esta sucesión es esencial, diciéndonos dónde el número irracional encaja entre los racionales.

Así, dos encajes de intervalos racionales $[a_n, b_n]$ y $[a_n', b_n']$ se denominan *equivalentes* si dan lugar a la misma separación de los números racionales en tres clases. El lector deberá probar como ejercicio que es necesario y suficiente para la equivalencia que: $a_n' - a_n$ sea una sucesión nula, o también: que las desigualdades

$$a_n \leq b_n', \qquad a_n' \leq b_n$$

se cumplan para todo n.

A un *encaje de intervalos racionales* $[a_n, b_n]$ se le asigna un *número real. Los números reales determinados por dos encajes distintos se considerarán iguales si las sucesiones son equivalentes.* Un número real está representado entonces por la separación de los números racionales en tres clases generada por encajes equivalentes de intervalos racionales. Si la segunda clase consiste de un número racional r, se considera al número real representado por esta separación en clases como idéntico con el número racional r.

[1] Llamada "Cortadura de Dedekind".

*c. Orden, límites y operaciones aritméticas para números reales

Habiendo definido los números reales, pueden definirse ahora las nociones de orden, suma, diferencia, producto, límite, etc., para éstos y probar que poseen las propiedades usuales. Para ser lógicamente coherentes, cualquier definición concerniente a números reales deberá: (1) poseer el significado ordinario en caso de que los números reales sean racionales, y (2) ser independiente de los encajes individuales de intervalos utilizados para representar números reales.

**Intervalos con puntos extremos reales*

Si bien hasta ahora, aún para la definición de números irracionales, los puntos extremos de los intervalos anidados fueron supuestos racionales, deberán quitarse aquí tales restricciones y mostrar que se puede operar con números reales exactamente como se hace con números racionales.

Los números reales se denotarán por letras $x, y, \ldots$. Si el número real x está dado por el encaje de intervalos racionales $[a_n, b_n]$ se escribe $x \sim \{[a_n, b_n]\}$. De la definición de número real se infiere una definición natural de orden para un número real, $x \sim \{[a_n, b_n]\}$, relativa a un número racional r. Se dice que $r < x$, $r = x$, $r > x$, según que r pertenezca a la primera, segunda, o tercera clase de la separación de los números racionales generada por el encaje de intervalos. Esta definición es obviamente independiente del encaje especial $\{[a_n, b_n]\}$ que define a x, y posee el significado ordinario cuando x es racional. Equivalentemente, se dice que $r < x$ si $r < a_n$ para casi todo n, $r = x$ si $a_n \leq r \leq b_n$ para todo n, y $r > x$ si $r > b_n$ para casi todo n.

Mediante la comparación de números reales con números racionales pueden compararse los números reales el uno con el otro. Sean $x \sim \{[a_n, b_n]\}$, $y \sim \{[\alpha_n, \beta_n]\}$. Se dice que $x < y$ si existe un número racional r tal que $x < r < y$. Claramente, esta definición no depende de las representaciones particulares de x e y mediante encajes puesto que las comparaciones con r racional son independientes de tales representaciones. Así, se dice que $x < y$ si existe un número racional r tal que $b_n < r < \alpha_n$ para casi todo n, o sea, simplemente, si $b_n < \alpha_n$ para casi todo n. La relación $x < y$ impide la posibilidad de que $y < x$ o $x = y$. Obviamente, $x < y$ e $y < z$ implican $x < z$.

Para dos números reales cualesquiera, x, y, se debe cumplir una de las relaciones $x < y$, $x = y$, $y < x$. Pues si $x \neq y$ y uno de ellos, por ejemplo, y, es racional, entonces y deberá estar en la primera o bien en la tercera clase de la separación generada por x, esto es, o bien es $y < x$ o es $x < y$. Si ni x ni y son racionales, las segundas clases de las correspondientes subdivisiones son vacías, y deberá existir un número racional r en la primera

clase respecto a uno de los números y en la tercera clase respecto al otro. Así se cumple una u otra de las relaciones $x < y$ o $y < x$.

Densidad. Una consecuencia inmediata de estas definiciones es la densidad de los números racionales, en el sentido de que entre dos números reales cualesquiera, x, y, existe siempre un número racional r. Se observa también que si un número real x está representado por un encaje de intervalos racionales $[a_n, b_n]$, entonces $a_n \leq x \leq b_n$ para todo n. Pues si $x < a_m$ para algún m, entonces $b_n < a_m$ para casi todo n, contradiciendo la desigualdad $a_m < b_n$, la cual se cumple para todo n. Por lo tanto, todo número real puede ser confinado en un intervalo racional $[a_n, b_n]$ de longitud arbitrariamente pequeña.

Una vez que los números reales están ordenados, puede hablarse de intervalos con *puntos extremos reales*. La densidad de los números racionales garantiza que todo tal intervalo incluye números racionales.

Límites. Un número real x se llama *límite* de una sucesión x_1, x_2, ... de números reales si todo intervalo abierto con puntos extremos reales que contiene a x contiene también elementos x_n para casi todo n. Esta definición es compatible con la definición en términos de intervalos racionales dada anteriormente, en el sentido de que un límite racional de una sucesión racional es un límite de la misma sucesión de números en el sentido más general de límite real. Como consecuencia de la definición de límite se encuentra que para un número real x representado por un encaje de intervalos racionales $[a_n, b_n]$ es

$$x = \lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n.$$

**Aritmética.* En seguida se definen las *operaciones aritméticas* para números reales $x \sim \{[a_n, b_n]\}$ y $y \sim \{[\alpha_n, {}_n\beta]\}$: Esto se logra más fácilmente para las operaciones de adición y substracción. Se define

$$x + y \sim \{[a_n + \alpha_n, b_n + \beta_n]\}, \qquad x - y \sim \{[a_n - \beta_n, b_n - \alpha_n]\}.$$

La prueba de que estas definiciones tienen sentido es un ejercicio sencillo cuyos detalles se dejan al lector (ver problema 3, p. 139). Por ejemplo, para $x - y$ es necesario verificar solamente que los intervalos $[a_n - \beta_n, b_n - \alpha]$ forman un encaje cuyas longitudes tienden a cero, y por lo tanto que ellos representan un número real z. El hecho de que z no dependa de las representaciones especiales de x e y se demuestra caracterizando la separación de números racionales en tres clases generada por z directamente en términos de x e y; por ejemplo, la primera clase consiste de los números racionales $r < z$, o sea de aquellos r que son excedidos por $a_n - \beta_n$ para algún n; y se ve fácilmente que estos r son números racionales de la forma $s - t$, donde s y t son números racionales para los cuales $s < x$ y $t > y$.

El producto de los dos números reales, x, y, está definido para $y > 0$ por

$$x \cdot y \sim \{[a_n\alpha_n, b_n\beta_n]\},$$

donde se ha supuesto que todo $\alpha_n > 0$; es obvio que los encajes son adecuados para utilizarse con xy en los casos $y < 0$ e $y = 0$. Siempre que y sea un número *racional* positivo, el producto $x \cdot y$ es representable también en la forma

$$x \cdot y \sim \{[a_n y, b_n y]\}.$$

Para un número natural $y = m$ el producto $x \cdot y = mx$ puede ser obtenido también por adición repetida de x, esto es, $mx = x + (m - 1)x = x + x + \cdots + x$.

Las operaciones aritméticas obedecen las reglas usuales. En particular, la relación $x < y$ es equivalente a $0 < y - x$. Se puede introducir el *valor absoluto* de un número real y probar la desigualdad del triángulo $|x + y| \leq |x| + |y|$. A la noción de límite de una sucesión de números reales, definida anteriormente en términos de relaciones de orden, puede dársele entonces la formulación equivalente: $x = \lim_{n \to \infty} x_n$ si para todo real positivo ε la relación $|x - x_n| < \varepsilon$ se cumple para casi todo n.

Verificaremos ahora el así llamado

AXIOMA DE ARQUÍMEDES. *Si x e y son números reales y x es positivo, existe entonces un número natural m tal que $mx > y$.*

En esencia, esto significa que un número real no puede ser "infinitamente pequeño" o bien 'infinitamente grande" comparado con otro (excepto si uno de ellos es cero). Para probar el axioma de Arquímedes (el cual en este contexto es realmente un *teorema)* se observa que para números racionales es una consecuencia de las propiedades comunes de los enteros. Si ahora $x \sim \{[a_n, b_n]\}$ e $y \sim \{[\alpha_n, \beta_n]\}$ son números reales y x es positivo, entonces $a_n > 0$ para casi todo n. Puesto que a_n y β_n son números racionales, puede entonces encontrarse un m tan grande que $ma_n > \beta_n$, de donde se sigue que $mx > \beta_n \geq y$.

d. Plenitud del continuo de números. Compacidad de intervalos cerrados. Criterios de convergencia

Los números reales hacen posible las operaciones límite con números racionales, pero serían de poco valor si las correspondientes operaciones de límite con los reales requirieran la introducción de algún tipo adicional de números "no reales" que habrían de ser intercalados entre los reales, y así sucesivamente *ad infinitum*. Afortunadamente, la definición de número real es tan amplia que no es posible ninguna extensión adicional del siste-

ma numérico sin deponer algunas de sus propiedades esenciales (tal como el "orden", que debe ser descartado para números complejos).

Principio de continuidad

Esta completitud del continuo de números reales está expresada por el principio básico de continuidad (ver p. 32); todo encaje de intervalos con puntos extremos reales contiene un número real. Para probar esto considérense los intervalos cerrados $[x_n, y_n]$, siendo cada intervalo contenido en el precedente, cuyas longitudes $y_n - x_n$ forman una sucesión nula. Se afirma que existe un número real x contenido en todo $[x_n, y_n]$: Las sucesiones x_n e y_n tendrán entonces como límite a x. Para probar esto se reemplaza la sucesión anidada $[x_n, y_n]$ por un encaje de intervalos racionales $[a_n, b_n]$ que contiene a los $[x_n, y_n]$. Esta sucesión o encaje racional definirá entonces el número real x deseado. Para cada n, sea a_n el mayor número racional de la forma $p/2^n$ menor que x_n, y b_n el menor número racional de la forma $q/2^n$ mayor que y_n, donde p y q son enteros. Claramente, los intervalos $[a_n, b_n]$ forman un encaje que representa un número real x. Si x estuviera situado fuera de uno de los intervalos $[x_m, y_m]$, digamos, $x < x_m$, existiría un racional r con $x < r < x_m$, de donde, para todo n suficientemente grande, se tendría

$$y_n \leq b_n < r < x_m \leq x_n,$$

lo cual es imposible. Por lo tanto, todos los intervalos $[x_m, y_m]$ contienen al punto x.

Principio de Weierstrass—Compacidad

Son importantes otras diversas versiones de este principio de *continuidad.* La primera es el principio de Weierstrass de *existencia de puntos límite o puntos de acumulación de sucesiones acotadas.* Un punto x es un *punto límite* de una sucesión $x_1, x_2, \ldots$ si todo intervalo abierto que contiene a x contiene también puntos x_n para una *infinidad de números naturales n.* Nótese la diferencia entre esta definición y la definición de *límite,* donde los x_n deben estar situados en el intervalo abierto para *casi todo n,* o sea, para todo n excepto a lo más un número finito, o para todo n suficientemente grande. Si una sucesión tiene un límite, entonces este límite es también un punto límite de la sucesión y es, de hecho, el único. Puede no haber punto límite (como en el ejemplo de la sucesión 1, 2, 3, 4, . . .) o bien un solo punto límite (como en una sucesión convergente) o diversos puntos límite (por ejemplo, la sucesión $1, -1, 1, -1, \ldots$ tiene los dos puntos límite $+1$ y -1). El principio de Weierstrass asegura que: *Toda sucesión acotada tiene al menos un punto límite.*

Para probar esto se observa que, como la sucesión $x_1, x_2, \ldots$ es *acotada,* existe un intervalo $[y_1, z_1]$ que contiene a todo x_n. Partiendo con $[y_1, z_1]$ se

construye por inducción sobre n un encaje de intervalos $[y_n, z_n]$ conteniendo cada uno puntos x_m para una infinidad de naturales m. Si $[y_n, z_n]$ contiene una infinidad de x_m, se divide $[y_n, z_n]$ en dos partes iguales por su punto medio. Por lo menos uno de los dos intervalos cerrados resultantes deberá contener, a su vez, una infinidad de x_m, y puede ser tomado como el intervalo $[y_{n+1}, z_{n+1}]$. Es claro que los $[y_n, z_n]$ forman un encaje que representa un número real x. Todo intervalo abierto que contiene a x contendrá los intervalos $[y_n, z_n]$ para n suficientemente grande y, por lo tanto, deberá contener una infinidad de x_m.

Los puntos límite pueden también ser definidos como *límites de subsucesiones* de la sucesión infinita $x_1, x_2, \ldots$ dada. Una subsucesión es cualquier sucesión infinita extraída de la sucesión dada, o sea, de la forma $x_{n_1}, x_{n_2}, x_{n_3}, \ldots$, donde $n_1 < n_2 < n_3 < \cdots$. Obviamente, un punto x es un punto límite de la sucesión $x_1, x_2, \ldots$ si es límite de alguna subsucesión. Recíprocamente, para todo punto límite x puede construirse por inducción una subsucesión $x_{n_1}, x_{n_2}, \ldots$ convergente a x. Si $x_{n_1}, \ldots, x_{n_{k-1}}$ están ya definidos, se toma para n_k un número de la infinidad de enteros n para los cuales $n > n_{k-1}$ y $|x_n - x| < 2^{-k}$.

Se reestablece el principio de Weierstrass en la forma:

TEOREMA. *Toda sucesión infinita acotada de números reales tiene una subsucesión convergente.*

Un conjunto se dice *compacto* si toda sucesión de sus elementos contiene una subsucesión convergente a un elemento del conjunto. En otras palabras, el teorema dice que *intervalos cerrados de números reales son conjuntos compactos.*

Sucesiones monótonas

Una consecuencia especial de este teorema es que *toda sucesión monótona acotada converge.* Así, sea la sucesión $x_1, x_2, \ldots$ monótona, por ejemplo, monótona creciente. Si la sucesión es también acotada, tiene un punto límite x. Arbitrariamente cerca de x deben existir puntos x_n de la sucesión, sin exceder ninguno a x puesto que los términos subsiguientes crecen y si $x_n > x$ entonces $x_m \geq x_n > x$ para $m > n$. Se sigue que todo intervalo que contiene a x contiene a casi todo x_n, o sea, que x es el límite de la sucesión.

Criterio de convergencia de Cauchy

La condición de que una sucesión sea acotada y monótona es *suficiente* para la convergencia. La importancia de este aserto es que permite a menudo probar la existencia del límite de una sucesión sin requerir un conocimiento *a priori* del valor del límite; además, la acotación y monotonía de una sucesión son propiedades que usualmente son fáciles de verificar en

aplicaciones concretas. Sin embargo, no toda sucesión convergente debe ser monótona (aunque debe ser acotada), y es importante disponer de un criterio de convergencia de aplicabilidad más general. Tal es la ***prueba intrínseca o criterio de convergencia de Cauchy,*** que es una condición necesaria y suficiente para la existencia del límite de una sucesión.

La sucesión $x_1, x_2, x_3, \ldots$ converge si y sólo si para todo ε positivo existe un N tal que $|x_n - x_m| < \varepsilon$ para todos n y m mayores que N.

En otras palabras, una sucesión converge si dos cualesquiera de sus elementos con índices suficientemente grandes difieren en menos de ε uno del otro.

Se afirma que la condición es necesaria para la convergencia. Si $x = \lim x_n$, entonces todo x_n con n suficientemente grande difiere de x en menos de $\varepsilon/2$ y, en consecuencia, por la desigualdad del triángulo, cualesquiera dos de tales valores x_n y x_m diferirán uno del otro en menos de ε. Recíprocamente, considérese una sucesión para la cual $|x_n - x_m| < \varepsilon$ para todo $\varepsilon > 0$, y para todos n, m suficientemente grandes. Existe entonces un valor N tal que casi todo x_n difiere de x_N en menos de 1. Esto significa que casi todo x_n puede ser encerrado en un intervalo de longitud 2. Puede entonces encontrarse un intervalo tan grande que contenga también el número finito de los x_n que pueden encontrarse fuera del intervalo en torno de x_N. Así, la sucesión está acotada y posee por lo tanto un punto límite x. Todo intervalo abierto que contiene a x contendrá también algunos puntos x_m con m arbitrariamente grande. Puesto que los puntos x_n difieren arbitrariamente poco uno del otro para n suficientemente grande, se sigue que el intervalo abierto en torno de x deberá contener a casi todo x_n, y así x resulta ser el límite de la sucesión.

e. Cotas. Extremo superior y extremo inferior *

Es de gran importancia que un conjunto acotado de números reales posea "las mejores cotas posibles" superiores e inferiores. Un conjunto S de números reales x es *acotado* si todos los números de S pueden ser encerrados en un mismo intervalo finito. Existen entonces *cotas superiores* de S, números B que no son excedidos por ningún número x de S:

$$x \leq B \qquad \text{para todo } x \text{ en } S.$$

Similarmente, existen *cotas inferiores* A de S:

$$A \leq x \qquad \text{para todo } x \text{ en } S.$$

* (N. del T.) En la América Latina algunos dicen "mínima cota superior" y "máxima cota inferior", por traducir del inglés "least upper bound" y "greatest lower bound", respectivamente.

Así, para el conjunto de recíprocos de números naturales, 1, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, . . . , cualquier número $B \geq 1$ es una cota superior y todo número $A \leq 0$, una cota inferior. Aquí el número 1, un miembro del conjunto, es la menor de las cotas superiores; y el número 0, punto límite de los elementos del conjunto, aunque no un miembro de él, es la mayor de las cotas inferiores. La menor de las cotas superiores de un conjunto de números reales se denomina a menudo supremo; la mayor de las cotas inferiores, su *infimum*. En general el *supremum* y el *infimum* de un conjunto son o bien miembros del conjunto o al menos límites de sucesiones de miembros del conjunto. En efecto, si la menor de las cotas superiores, b de S, no pertenece a S, deberán existir algunos miembros de S situados arbitrariamente cerca de b, pues de otra forma podrían encontrarse cotas superiores de S menores que b; así, pueden seleccionarse sucesivamente una sucesión de números $x_1, x_2, \ldots$ de S situados cada vez más cerca de b y convergentes a b.

La *existencia de la menor de las cotas superiores* de un conjunto acotado S se sigue de inmediato de la convergencia de sucesiones monótonas acotadas. Para cualquier n defínase B_n como la menor cota racional superior de S con denominador 2^n. Claramente, para cualquier x en S y cualquier n

$$x \leq B_{n+1} \leq B_n \leq B_1.$$

Por lo tanto, los B_n forman una sucesión monótona decreciente y acotada la cual debe tener un límite b. Es fácil ver que b es una cota superior de S y que no existe ninguna cota superior menor. La existencia de la máxima cota inferior se prueba de la misma manera.

f. Numerabilidad de los números racionales

A fines del siglo XIX fue hecho un sorprendente descubrimiento relativo a los números racionales, que estimuló la creación, por Georg Cantor, de la Teoría de Conjuntos después de 1872. Aun cuando los números racionales son densos y no pueden ser ordenados por tamaño, pueden, sin embargo, ser arreglados como una sucesión infinita $r_1, r_2, \ldots, r_n, \ldots$ en la cual todo número racional aparece una vez. De esta forma los números racionales pueden ser numerados, o sea, ser contados como un primero, segundo, . . . , n-ésimo, . . . número racional, donde, por supuesto, el orden de los números en la sucesión no corresponde a su orden por magnitud. Este resultado, que se cumple también para números racionales en cualquier intervalo, es expresado por la propiedad: *Los números racionales son numerables, o sea, constituyen un conjunto numerable.*

Para probar este resultado se da simplemente una regla para arreglar los números racionales positivos como una sucesión. Cada uno de tales números puede ser escrito en la forma p/q, donde p y q son números natu-

rales. Para cada entero positivo k existen exactamente $k - 1$ fracciones p/q para las cuales $p + q = k$. Estas son arregladas en orden creciente de p. Escribiendo los diferentes arreglos de números para $k = 2, 3, 4, \ldots$ sucesivamente, se obtiene una sucesión que contiene todos los números racionales positivos. Omitiendo las fracciones en las cuales el numerador y el denominador tienen un factor común mayor que 1, y por lo tanto representan el mismo número racional que una fracción anterior, se obtiene la sucesión

$$\tfrac{1}{1}, \tfrac{1}{2}, \tfrac{2}{1}, \tfrac{1}{3}, \tfrac{3}{1}, \tfrac{1}{4}, \tfrac{2}{3}, \tfrac{3}{2}, \tfrac{4}{1}, \tfrac{1}{5}, \tfrac{5}{1}, \tfrac{1}{6}, \tfrac{2}{5}, \ldots$$

en la cual todo número racional positivo aparece exactamente una vez. Una sucesión análoga conteniendo a todos los números racionales, o bien, todos los números racionales en algún intervalo particular, es construida fácilmente.

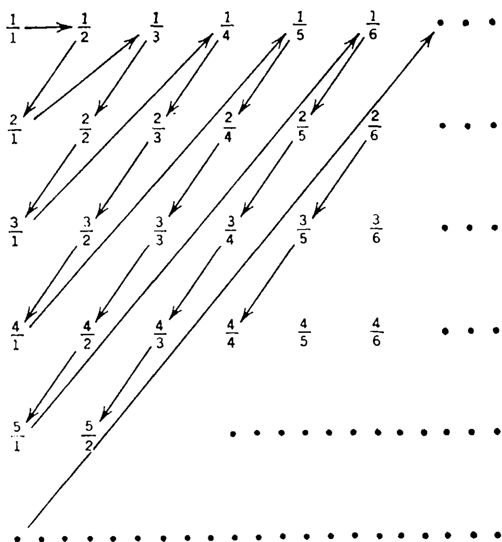


Figura 1.S.1. Numerabilidad de los racionales positivos.

Este resultado es visto en su debida perspectiva sólo a la luz de otro hecho básico: que *el conjunto de todos los* números *reales* no es numerable.[1] Esta es una indicación de que el conjunto de números reales contiene "muchos más" elementos que el de los números racionales, aun cuando ambos conjuntos son infinitos. Así, la numerabilidad es en realidad una propiedad altamente restrictiva de un conjunto.

[1] Para la demostración y una breve discusión general de los hechos básicos de la teoría de conjuntos, véase *What Is Mathematics?* por Courant y Robbins, p. 81.

La Teoría de Conjuntos desempeña un importante papel esclarecedor en matemáticas, aunque su uso con libre generalidad ha dado lugar a resultados paradójicos y controversias. Tales paradojas, sin embargo, no afectan la esencia de las matemáticas constructivas y están ausentes en la teoría de conjuntos para números reales.

S.2 Teoremas sobre funciones continuas

Propiedades importantes de las funciones continuas son establecidas sobre la base de la propiedad de completitud de los números reales. Recuérdese la definición de continuidad: la función $f(x)$ es continua en el punto ξ si para todo ε positivo dado la desigualdad $|f(x) - f(\xi)| < \varepsilon$ se cumple para todo x que difiere de ξ en menos de una cantidad adecuada δ, la cual depende generalmente de la elección de ε y ξ. Se sobreentiende en esta definición que se consideran solamente valores de x y ξ para los cuales f está definida.

Una definición más concisa de continuidad en términos de convergencia de sucesiones es la siguiente: *$f(x)$ es continua en el punto ξ si* $\lim_{n\to\infty} f(x_n) = f(\xi)$ *para toda sucesión $x_1, x_2, \ldots$ con límite ξ* (donde, nuevamente, los valores x_n y ξ están en el dominio de f). La equivalencia de las dos definiciones se probó en la sección 1.8, p. 105.

La función f se denomina *continua en un intervalo* si f es continua en cada punto del intervalo. La función $f(x)$ es *uniformemente continua* si para $\varepsilon > 0$ dado se tiene $|f(x) - f(\xi)| < \varepsilon$ siempre que x y ξ se encuentren suficientemente cercanos, independientemente de su posición en el intervalo; así, f es uniformemente continua si la cantidad δ que aparece en la definición de continuidad puede ser elegida independiente de ξ: Para todo $\varepsilon > 0$ existe un $\delta = \delta(\varepsilon) > 0$ tal que $|f(x) - f(\xi)| < \varepsilon$ siempre que $|x - \xi| < \delta$. Para propósitos prácticos, esto significa que si se subdivide el intervalo en el cual f está definida en un número suficientemente grande de intervalos iguales, entonces f variará en menos que una cantidad prescrita ε en cada subintervalo. En cualquier punto f diferirá entonces en menos de ε de su valor en cualquier otro punto del mismo subintervalo.

Se probará ahora lo siguiente: *Toda función continua en un intervalo cerrado $[a, b]$ es uniformemente continua en ese intervalo.*

Si f no fuese uniformemente continua en $[a, b]$ existiría un $\varepsilon > 0$ fijo y puntos, x, ξ, en $[a, b]$, arbitrariamente cercanos uno del otro, para los cuales $|f(x) - f(\xi)| \geq \varepsilon$. Entonces sería posible para cualquier n elegir puntos x_n, ξ_n en $[a, b]$ para los cuales $|f(x_n) - f(\xi_n)| \geq \varepsilon$ y $|x_n - \xi_n| < 1/n$. Puesto que los x_n forman una sucesión acotada de números, podría encontrarse una subsucesión convergente a un punto η del intervalo (utilizando la *compacidad* de intervalos cerrados). Los correspondien-

tes valores ξ_n entonces convergerían también a η: puesto que f es continua en η se encontraría que $\eta = \lim f(x_n) = \lim f(\xi_n)$ para n tendiendo a infinito en la subsucesión, lo cual es imposible si $|f(x_n) - f(\xi_n)| \geq \varepsilon$ para todo n.

El *teorema del valor intermedio* afirma que: Si para una función $f(x)$ continua en un intervalo $a \leq x \leq b$, γ es cualquier valor entre $f(a)$ y $f(b)$, entonces $f(\xi) = \gamma$ para algún ξ adecuado entre a y b. Así, la *existencia de una solución* ξ de la ecuación $f(\xi) = \gamma$ es cierta si se exhiben dos valores a y b para los cuales $f(a) < \gamma$ y $f(b) > \gamma$ respectivamente. Esto implica inmediatamente la existencia de una *función inversa* determinada de manera única si f es continua y monótona, como se ha visto (pp. 67-68).

Para probar el teorema del valor intermedio, sean $a < b$, $f(a) = \alpha$, $f(b) = \beta$ y $\alpha < \gamma < \beta$. Sea S el conjunto de puntos x del intervalo $[a, b]$ para los cuales $f(x) < \gamma$. S es acotado y posee una misma cota superior ξ que pertenece también al intervalo cerrado $[a, b]$. Entonces $f(x) \geq \gamma$ para $\xi < x \leq b$. El punto ξ pertenece a S o bien es el límite de una sucesión de puntos x_n de S. En el primer caso, $f(\xi) < \gamma$; por lo tanto $\xi < b$, puesto que $f(b) > \gamma$, y existen puntos x entre ξ y b, arbitrariamente cerca de ξ, para los cuales $f(x) \geq \gamma$. Esto es imposible si f es continua en ξ y $f(\xi) < \gamma$. En el segundo caso, $f(\xi) \geq \gamma$, y se encuentra de $f(x_n) < \gamma$ y $\lim_{n \to \infty} x_n = \xi$ que $f(\xi) \leq \gamma$; y como se vio ya que $f(\xi) < \gamma$ es imposible, se debe tener $f(\xi) = \gamma$.

Una tercera propiedad básica de una función continua $f(x)$ en un intervalo cerrado $[a, b]$ es la *existencia de un valor máximo (máximo)*, entendiéndose que existe un punto ξ en el intervalo $[a, b]$ tal que $f(x) \leq f(\xi)$ para todo x en el intervalo. Análogamente, f alcanzará su *valor mínimo (mínimo)* en algún punto η del intervalo: $f(x) \geq f(\eta)$ para todo x en el intervalo. Es esencial que el intervalo sea *cerrado:* por ejemplo, las funciones $f(x) = x$, $f(x) = 1/x$ son continuas pero no poseen un valor máximo en el intervalo abierto $0 < x < 1$; el máximo puede ocurrir justamente en uno de los puntos extremos o bien no existir si f no es continua en estos puntos.

Para probar este principio se observa que una función f continua en $[a, b]$ es necesariamente *acotada:* esto es, los valores $f(x)$ que constituyen el "rango" S de f están situados en algún intervalo finito. Realmente, por la continuidad uniforme de f pueden encontrarse un número finito de puntos $x_1, x_2, \ldots, x_n$ en el intervalo tales que $f(x)$ en cualquier x del intervalo difiera en menos de uno de alguno de los números $f(x_1), f(x_2), \ldots, f(x_n)$, los cuales pueden ser todos encajados en un intervalo finito. Puesto que entonces el conjunto S de valores $f(x)$ es acotado, posee una mínima cota superior M. Este M es el menor número tal que $f(x) \leq M$ para todo x en

$[a, b]$. M pertenece a S o bien es el límite de una sucesión de puntos de S. En el primer caso existe un ξ en $[a, b]$ tal que $f(\xi) = M$. En el segundo caso existe una sucesión de puntos x_n en $[a, b]$ tal que $\lim_{n \to \infty} f(x_n) = M$; así, puede encontrarse una subsucesión de los x_n que converge a un punto ξ de $[a, b]$, y entonces $f(\xi) = M$ por continuidad de f en ξ. Claramente, $f(\xi)$ es el máximo de f.

S.3 Coordenadas polares

En el capítulo 1 se han representado funciones geométricamente por curvas. La geometría analítica sigue el procedimiento inverso, principiando con una curva y representándola mediante una función, por ejemplo, mediante una función que expresa una de las coordenadas de un punto de la curva en términos de la otra. Este punto de vista nos conduce naturalmente a considerar, además de las coordenadas rectangulares a las cuales nos hemos restringido, otros sistemas de coordenadas posiblemente mejor situados para la representación de curvas dadas geométricamente. El ejemplo más importante es el de *coordenadas polares* r, θ conectadas con las coordenadas rectangulares x, y de un punto P por las ecuaciones

$$x = r \cos \theta, \quad y = r \operatorname{sen} \theta, \quad r^2 = x^2 + y^2, \quad \tan \theta = \frac{y}{x},$$

cuya interpretación geométrica se hace clara en la Fig. 1.S.2.[1]

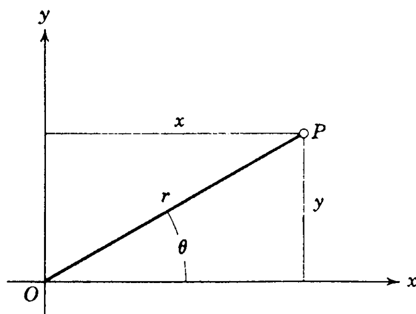


Figura 1.S.2 Coordenadas polares.

Considérese, por ejemplo, el *lemniscato*. Esta curva está geométricamente definida como el lugar geométrico de todos los puntos P para los cuales el producto de sus distancias r_1 y r_2 a los puntos fijos F_1 y F_2 cuyas

[1] Las coordenadas polares no están determinadas totalmente por el punto P. Además de θ, cualquiera de los ángulos $\theta \pm 2\pi$, $\theta \pm 4\pi$, ... puede ser considerado un ángulo polar de P.

coordenadas rectangulares son $x = a$, $y = 0$ y $x = -a$, $y = 0$ respectivamente, tiene el valor constante a^2 (ver Fig. 1.S.3). Puesto que

$$r_1^2 = (x - a)^2 + y^2, \quad r_2^2 = (x + a)^2 + y^2,$$

mediante un cálculo sencillo se obtiene la ecuación del lemniscato en la forma

$$(x^2 + y^2)^2 - 2a^2(x^2 - y^2) = 0.$$

Introduciendo coordenadas polares se obtiene

$$r^4 - 2a^2r^2(\cos^2\theta - \operatorname{sen}^2\theta) = 0;$$

dividiendo por r^2 y utilizando una fórmula trigonométrica sencilla obtenemos finalmente

$$r^2 = 2a^2 \cos 2\theta.$$

Así, la ecuación del lemniscato es más sencilla en coordenadas polares que en coordenadas rectangulares.

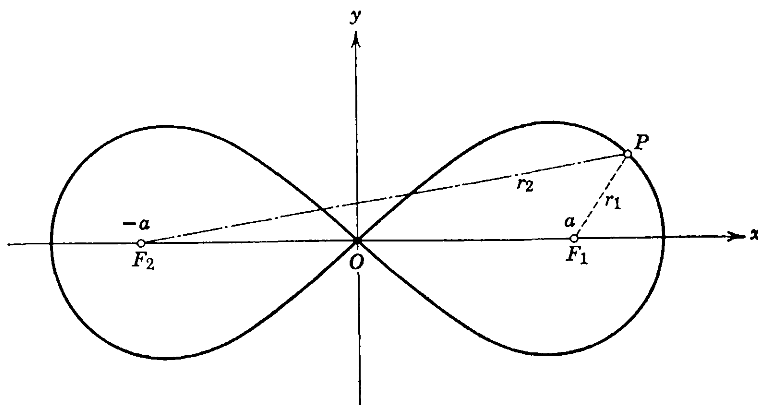


Figura 1.S.3 Lemniscato.

S.4 Observaciones sobre los números complejos

Nuestros estudios se basarán principalmente en el continuo de los números reales. Sin embargo, con miras hacia las discusiones de los capítulos 7, 8 y 9, se recuerda al lector que los problemas del álgebra han conducido hacia una extensión aún más amplia del concepto de número, los *números complejos*. El paso de los números naturales a los números reales surgió del deseo de eliminar fenómenos excepcionales y hacer siempre posible ciertas operaciones, tales como la substracción, la división y la correspondencia entre puntos y números. Análogamente, estamos precisados a introducir los números complejos, por el requisito de que toda ecuación cuadrática, y,

de hecho, toda ecuación algebraica, deberá tener solución. Si, por ejemplo, se desea que la ecuación

$$x^2 + 1 = 0$$

tenga raíces, estamos obligados a introducir nuevos símbolos i y $-i$ como sus raíces. (Como se muestra en la teoría de funciones de variable compleja, esto es suficiente para asegurar que *toda* ecuación algebraica tenga una solución.)[1]

Si a y b son números reales ordinarios, el *número complejo* $c = a + ib$ denota una pareja de números (a, b) con la cual se efectúan los cálculos de acuerdo a la siguiente regla general: Los números complejos (entre los cuales los números reales están incluidos como el caso especial $b = 0$) se suman, multiplican y dividen, tratándose al símbolo i como una cantidad indeterminada y simplificando todas las expresiones haciendo uso de la ecuación $i^2 = -1$ para eliminar todas las potencias de i mayores que la primera, dejando solamente una expresión de la forma $a + ib$.

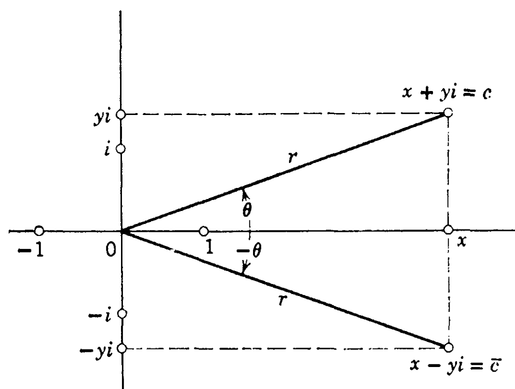


Figura 1.S.4 Representación geométrica de un número complejo $x + yi$ y de su conjugado.

Se supone que el lector ya posee un cierto grado de familiaridad con los números complejos. Se recalcará, sin embargo, una relación particularmente importante que será explicada en conexión con la representación geométrica o trigonométrica de los números complejos. Si $c = x + iy$ es uno de tales números, éste se representa en un sistema coordenado rectangular mediante el punto P con coordenadas x e y. Por medio de las ecuaciones $x = r \cos \theta$, $y = r \operatorname{sen} \theta$, se introducen las coordenadas r y θ (ver p. 125) en lugar de las coordenadas rectangulares x e y. Entonces $r = \sqrt{x^2 + y^2}$ es la distancia del punto P al origen y θ, el ángulo entre

[1] Una ecuación algebraica es de la forma $P(x) = 0$, donde P es un polinomio con coeficientes complejos.

el eje x positivo y el segmento OP. El número complejo c está representado en la forma

$$c = r(\cos\theta + i \operatorname{sen}\theta).$$

El ángulo θ se denomina la *amplitud* del número complejo c; la cantidad r, su *valor absoluto* o *módulo,* para el cual se escribe también $|c|$. Al número complejo "conjugado" $\bar{c} = x - iy$ obviamente le corresponde el mismo valor absoluto pero la amplitud $-\theta$ (Fig. 1.S.4). Claramente,

$$r^2 = |c|^2 = c\bar{c} = x^2 + y^2.$$

Si se utiliza esta representación trigonométrica, la multiplicación de números complejos toma una forma particularmente simple, pues entonces

$$\begin{aligned} c \cdot c' &= r(\cos\theta + i \operatorname{sen}\theta) \cdot r'(\cos\theta' + i \operatorname{sen}\theta') \\ &= rr'[(\cos\theta\cos\theta' - \operatorname{sen}\theta \operatorname{sen}\theta') + i(\cos\theta \operatorname{sen}\theta' + \operatorname{sen}\theta\cos\theta')]. \end{aligned}$$

Si se usan los teoremas de adición para las funciones trigonométricas, esto se transforma en

$$c \cdot c' = rr'(\cos(\theta + \theta') + i \operatorname{sen}(\theta + \theta')).$$

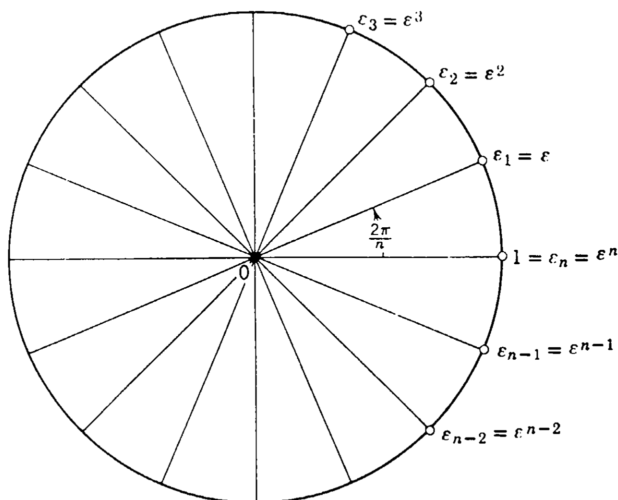


Figura 1.S.5 Las raíces n-ésimas de la unidad (para $n = 16$).

En consecuencia, los números complejos se multiplican multiplicando sus valores absolutos y sumando sus amplitudes. La notable fórmula

$$(\cos\theta + i \operatorname{sen}\theta)(\cos\theta' + i \operatorname{sen}\theta') = \cos(\theta + \theta') + i \operatorname{sen}(\theta + \theta')$$

se denomina usualmente *teorema de De Moivre.* Este nos conduce a la relación

$$(\cos\theta + i\,\text{sen}\,\theta)^n = \cos n\theta + i\,\text{sen}\,n\theta,$$

la cual, por ejemplo, permite de inmediato resolver la ecuación $x^n = 1$ para enteros positivos n; las raíces (las así llamadas raíces de la unidad) son

$$\varepsilon_1 = \varepsilon = \cos\frac{2\pi}{n} + i\,\text{sen}\,\frac{2\pi}{n}, \qquad \varepsilon_2 = \varepsilon^2 = \cos\frac{4\pi}{n} + i\,\text{sen}\,\frac{4\pi}{n}, \ldots,$$

$$\varepsilon_{n-1} = \varepsilon^{n-1} = \cos\frac{2(n-1)\pi}{n} + i\,\text{sen}\,\frac{2(n-1)\pi}{n}, \qquad \varepsilon_n = \varepsilon^n = 1$$

(Fig. 1.S.5).

Geométricamente, los puntos que corresponden a las raíces de la unidad forman los vértices de un n-ágono regular inscrito en el círculo de radio 1 con centro en el origen.

Finalmente, si nos imaginamos la expresión del miembro izquierdo de la ecuación $(\cos\theta + i\,\text{sen}\,\theta)^n = \cos n\theta + i\,\text{sen}\,n\theta$ desarrollada por el teorema binomial, necesitaremos solamente separar las partes real e imaginaria de manera de obtener expresiones para $\cos n\theta$ y $\text{sen}\,n\theta$ en términos de potencias y productos de potencias de $\text{sen}\,\theta$ y $\cos\theta$:

$$\cos n\theta = \cos^n\theta - \binom{n}{2}\cos^{n-2}\theta\,\text{sen}^2\,\theta + \binom{n}{4}\cos^{n-4}\theta\,\text{sen}^4\,\theta + - \cdots,$$

$$\text{sen}\,n\theta = \binom{n}{1}\cos^{n-1}\theta\,\text{sen}\,\theta - \binom{n}{3}\cos^{n-3}\theta\,\text{sen}^3\,\theta + \binom{n}{5}\cos^{n-5}\theta\,\text{sen}^5\,\theta + - \cdots.$$

PROBLEMAS

SECCION 1.1a, página 26

1. (*a*) Si a es racional y si x es irracional, demuéstrese que $a + x$ es irracional, y si $a \neq 0$, que ax es irracional.

(*b*) Muéstrese que entre dos números racionales cualesquiera existe por lo menos un número irracional y, en consecuencia, una infinidad de ellos.

2. Probar que los números siguientes no son racionales: (*a*) $\sqrt{3}$. (*b*) $\sqrt{n}$, donde el entero n no es un cuadrado perfecto, es decir, no es el cuadrado de un entero. (*c*) $\sqrt[3]{2}$. (*d*) $\sqrt[p]{n}$, donde n no es una potencia p-ésima perfecta.

***3.** (*a*) Para cualquier raíz racional de un polinomio con coeficientes enteros,

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0, \qquad (a_n \neq 0),$$

expresada en sus menores términos como p/q, demuéstrese que el numerador p es factor de a_0 y que el denominador q es factor de a_n. (Por medio de este criterio podemos obtener todas las raíces reales racionales y, por lo tanto, demostrar la irracionalidad de cualesquiera otras raíces reales.)

(*b*) Demuéstrese la irracionalidad de $\sqrt{2} + \sqrt[3]{2}$ y $\sqrt{3} + \sqrt[3]{2}$.

SECCION 1.1c, página 33

1. Denotemos por $[x]$ la parte entera de x; es decir, $[x]$ es el entero para el que se cumple

$$x - 1 < [x] \leq x.$$

Sean $c_0 = [x]$, y $c_n = [10^n(x - c_0) - 10^{n-1}c_1 - 10^{n-2}c_2 - \cdots - 10c_{n-1}]$ para $n = 1, 2, 3 \ldots$. Verifíquese que la representación decimal de x es

$$x = c_0 + 0 \cdot c_1c_2c_3 \cdots$$

y que en esta construcción queda excluida la posibilidad de obtener una cadena infinita de nueves.

2. Defínase la desigualdad $x > y$ para dos números reales en términos de sus representaciones decimales (véase el suplemento, pág. 115).

***3.** Si p y q son enteros, $q > 0$, demuéstrese que el desarrollo decimal de p/q o bien termina (de manera que todos los dígitos posteriores al último sean ceros) o es periódico; es decir, hay un punto a partir del que el desarrollo decimal consiste en la repetición sucesional de una determinada cadena de dígitos. Por ejemplo, tenemos que $\frac{1}{4} = 0.25$ termina, y que $1/11 = 0.090909 \cdots$ es periódico. La longitud de la cadena que se repite se llama *período* del decimal; para $1/11$ el período es 2. En general, ¿qué tan grande puede ser el período de p/q?

SECCION 1.1e, página 36

1. Por medio de sólo signos de desigualdad (sin emplear los signos de valor absoluto) especifíquense los valores de x que satisfacen las relaciones siguientes. Discútanse todos los casos.

(*a*) $|x - a| < |x - b|$.
(*b*) $|x - a| < x - b$.
(*c*) $|x^2 - a| < b$.

2. Un intervalo (véanse las definiciones del texto) puede ser definido como cualquier porción conexa del continuo real. Se dice que un subconjunto S del continuo real es *conexo* si con todo par de puntos a, b en S, el conjunto S contiene todo el intervalo cerrado $[a, b]$. Al margen de los intervalos abiertos y cerrados que hemos ya mencionado, existen los intervalos "semiabiertos" $a \leq x < b$ y $a < x \leq b$ (que a veces se denotan respectivamente por $[a, b)$ y $(a, b])$ y los intervalos no acotados que pueden constituir o bien toda la recta real, o bien un rayo, es decir, una "semirrecta" $x \leq a$, $x < a$, $x > a$, $x \geq a$ (que a veces se denotan respectivamente $(-\infty, \infty)]$ y $(-\infty, a]$, $(-\infty, a)$, (a, ∞), $[a, \infty))$, (véase también la nota al pie de la página, en la p. 46).

*(*a*) Demuéstrese que los casos de intervalos que se acaban de describir agotan todas las posibilidades de los subconjuntos conexos del eje numérico.

(*b*) Determínense los intervalos en los cuales se satisfacen las desigualdades siguientes.

(i) $x^2 - 3x + 2 < 0$.
(ii) $(x - a)(x - b)(x - c) > 0$, para $a < b < c$.
(iii) $|1 - x| - x \geq 0$.
(iv) $\dfrac{x - a}{x + a} \geq 0$.
(v) $\left|x + \dfrac{1}{x}\right| \geq 6$.

(vi) $[x] \leq x/2$. Véase el problema 1 de la sección 1.1c.

(vii) $\operatorname{sen} x \geq \sqrt{2}/2$.

(*c*) Demuéstrese que si $a \leq x \leq b$, entonces $|x| \leq |a| + |b|$.

3. Dedúzcanse las desigualdades

(*a*) $x + \dfrac{1}{x} \geq 2$, para $x > 0$,

(*b*) $x + \dfrac{1}{x} \leq -2$, para $x < 0$,

(*c*) $\quad + \left.\dfrac{1}{x}\right| \geq 2$, para $x \neq 0$.

4. La media armónica ξ de dos números positivos a, b se define por

$$\frac{1}{\xi} = \frac{1}{2}\left(\frac{1}{a} + \frac{1}{b}\right).$$

Demuéstrese que la media armónica no excede la media geométrica; es decir, que $\xi \leq \sqrt{ab}$. ¿Cuándo son iguales las dos medias?

5. Dedúzcanse las desigualdades siguientes:

(*a*) $x^2 + xy + y^2 \geq 0$,

*(*b*) $x^{2n} + x^{2n-1}y + x^{2n-2}y^2 + \cdots + y^{2n} \geq 0$,

*(*c*) $x^4 - 3x^3 + 4x^2 - 3x + 1 \geq 0$.

¿Cuándo se cumple la igualdad?

*6. ¿Cuál es la interpretación geométrica de la desigualdad de Cauchy para $n = 2, 3$?

7. Demuéstrese que la igualdad se cumple en la desigualdad de Cauchy si y sólo si las a_ν son proporcionales a las b_ν: es decir, $ca_\nu + db_\nu = 0$ para toda ν, donde c y d no dependen de ν y no son ambas cero.

8. (*a*) $|x - a_1| + |x - a_2| + |x - a_3| \geq a_3 - a_1$, para $a_1, < a_2 < a_3$.

¿Para qué valores de x se cumple la igualdad?

*(*b*) Encontrar el valor máximo de y para el que se tiene, para toda x, que

$$|x - a_1| + |x - a_2| + \cdots + |x - a_n| \geq y,$$

donde $a_1 < a_2 < \cdots < a_n$. ¿En qué condiciones se cumple la igualdad?

9. Demuéstrese que las desigualdades siguientes se cumplen cuando a, b, c son positivos.

(*a*) $a^2 + b^2 + c^2 \geq ab + bc + ca$.

(*b*) $(a + b)(b + c)(c + a) \geq 8abc$.

(*c*) $a^2b^2 + b^2c^2 + c^2a^2 \geq abc(a + b + c)$.

10. Supongamos que los números x_1, x_2, x_3 y a_{ik} $(i, k = 1, 2, 3)$ son todos positivos y además, $a_{ik} \leq M$ y $x_1{}^2 + x_2{}^2 + x_3{}^2 \leq 1$. Demuéstrese que

$$a_{11}x_1{}^2 + a_{12}x_1x_2 + \cdots + a_{33}x_3{}^2 \leq 3M.$$

*11. Demuéstrese la desigualdad siguiente e interprétese geométricamente para $n \leq 3$.

$$\sqrt{(a_1 - b_1)^2 + \cdots + (a_n - b_n)^2} \leq \sqrt{(a_1{}^2 + \cdots + a_n{}^2)}\ \sqrt{(b_1{}^2 + \cdots + b_n{}^2)}.$$

12. Demuéstrese e interprétese geométricamente que para $n \leq 3$,

$$\sqrt{(a_1 + b_1 + \cdots + z_1)^2 + \cdots + (a_n + b_n + \cdots + z_n)^2}$$
$$\leq \sqrt{a_1{}^2 + \cdots + a_n{}^2} + \sqrt{b_1{}^2 + \cdots + b_n{}^2} + \cdots + \sqrt{z_1{}^2 + \cdots + z_n{}^2}.$$

13. Demuéstrese que la media geométrica de n números positivos no es mayor que la media aritmética; es decir, si $a_i > 0$ $(i = 1, \ldots, n)$, entonces

$$\sqrt[n]{a_1a_2\cdots a_n} \leq \frac{1}{n}(a_1 + a_2 + \cdots + a_n).$$

(*Indicación:* supóngase que $a_1 \leq a_2 \leq \cdots \leq a_n$. Para el primer paso reemplácese a_n por la media geométrica y ajústese a_1 de manera que la media geométrica quede sin alteración.)

SECCION 1.2d, página 55

1. Si $f(x)$ es continua en $x = a$ y $f(a) > 0$, mostrar que el dominio de f contiene un intervalo abierto alrededor de a en el que $f(x) > 0$.

2. Por medio de la definición de continuidad muéstrese que los intervalos centrados

$$|f(x) - f(x_0)| < \varepsilon \quad \text{y} \quad |x - x_0| < \delta$$

se pueden substituir por un intervalo abierto arbitrario que contenga $f(x_0)$ y otro intervalo abierto suficientemente pequeño que contenga x_0, como se indica en la página 57.

3. Sea $f(x)$ continua para $0 \leq x \leq 1$. Supongamos además que $f(x)$ toma valores racionales y ningunos otros y que $f(x) = \frac{1}{2}$ cuando $x = \frac{1}{2}$. Demuéstrese que en todo punto $f(x) = \frac{1}{2}$.

4. (*a*) Definamos $f(x)$ para todos los valores de x de la manera siguiente:

$$f(x) = \begin{cases} 0, & x \text{ irracional} \\ 1, & x \text{ racional.} \end{cases}$$

Demuéstrese que $f(x)$ es discontinua en todo punto.

(*b*) Consideremos por otra parte

$$g(x) = \begin{cases} 0, & x \text{ irracional} \\ \dfrac{1}{q} & x = \dfrac{p}{q} \text{ racional en sus menores términos.} \end{cases}$$

(Se dice que el número racional p/q está expresado en sus menores términos si los enteros p y q no tienen factores comunes mayores que 1, y $q > 0$. De esta manera, $f(16/29) = 1/29$). Demuéstrese que $g(x)$ es continua para todos los valores irracionales y discontinua para todos los valores racionales.

***5.** Si $f(x)$ satisface la ecuación funcional

$$f(x + y) = f(x) + f(y)$$

para todos los valores de x e y, encuéntrense los valores de $f(x)$ para los valores racionales de x y demuéstrese, si $f(x)$ es continua, que $f(x) = cx$, donde c es constante.

6. (*a*) Si $f(x) = x^n$, encuéntrese una δ que pueda depender de ξ tal que

$$|f(x) - f(\xi)| < \varepsilon$$

siempre que

$$|x - \xi| < \delta.$$

*(*b*) Hágase lo mismo si $f(x)$ es un polinomio cualquiera

$$f(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0,$$

donde $a_n \neq 0$.

SECCION 1.2e, página 68

1. Demuéstrese que si $f(x)$ es monótona en $[a, b]$ y satisface la propiedad del valor intermedio, entonces $f(x)$ es continua. ¿Se puede extraer la misma conclusión si f no es monótona?

2. (*a*) Demuéstrese que x^n es monótona para $x > 0$. En consecuencia, descríbase para $a > 0$ que $x^n = a$ posee una solución positiva única $\sqrt[n]{a}$.

(*b*) Sea $f(x)$ un polinomio

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0, \qquad (a_n \neq 0).$$

Demuéstrese que (i) si n es impar entonces $f(x)$ tiene por lo menos una raíz real, (ii) si a_n y a_0 son de signos opuestos, entonces $f(x)$ tiene no menos de una raíz positiva y además si n es par, $n \neq 0$, entonces $f(x)$ tiene también una raíz negativa.

*3. (*a*) Demuéstrese que existe una recta en cada dirección tal que bisecta cualquier triángulo dado, es decir, tal que lo divide en dos porciones de áreas iguales.

(*b*) Demuéstrese para cualquier par de triángulos que existe una recta que los bisecta simultáneamente.

SECCION 1.3b, página 72

1. (*a*) Demuéstrese que $\sqrt{x}$ no es una función racional. (*Indicación:* examínese la posibilidad de representar $\sqrt{x}$ como función racional para $x = y^2$. (Puede servir el hecho de que un polinomio diferente de cero no puede tener más que un número finito de raíces.)

(*b*) Demuéstrese que $\sqrt[n]{x}$ no es una función racional.

SECCION 1.3c, página 73

1. (*a*) Muéstrese que una recta no puede cortar la gráfica de una función polinomial de grado mayor que uno sino en un número finito de puntos.

(*b*) Obténgase el mismo resultado para las funciones racionales en general.

(*c*) Verifíquese que las funciones trigonométricas no son racionales.

SECCION 1.5, página 80

1. Demuéstrense las propiedades siguientes de los coeficientes binomiales.

(*a*) $1 + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n-1} + \binom{n}{n} = 2^n.$

(*b*) $1 - \binom{n}{1} + \binom{n}{2} - \binom{n}{3} + \cdots + (-1)^n\binom{n}{n} = 0.$

(*c*) $\binom{n}{1} + 2\binom{n}{2} + 3\binom{n}{3} + \cdots + n\binom{n}{n} = n(2^{n-1}).$ (*Indicación:* represénten-se los coeficientes binomiales en términos de factoriales.)

(*d*) $1 \cdot 2\binom{n}{2} + 2 \cdot 3\binom{n}{3} + \cdots + (n-1)n\binom{n}{n} = n(n-1)2^{n-2}.$

(*e*) $1 + \frac{1}{2}\binom{n}{1} + \frac{1}{3}\binom{n}{2} + \cdots + \frac{1}{n+1}\binom{n}{n} = \frac{2^{n+1}-1}{n+1}.$

*(*f*) $\binom{n}{0}^2 + \binom{n}{1}^2 + \cdots + \binom{n}{n}^2 = \binom{2n}{n}.$ (*Indicación:* considérese el coeficiente de x^n en $(1 + x)^{2n}$.)

*(g) $$S_n = \binom{n}{0} - \frac{1}{3}\binom{n}{1} + \frac{1}{5}\binom{n}{2} - \frac{1}{7}\binom{n}{3} + \cdots + \frac{(-1)^n}{2n+1}\binom{n}{n}$$
$$= \frac{4^n(n!)^2}{(2_n+1)!}.$$

$\left(\right.$*Indicación:* demuéstrese que $\frac{2n+2}{2n+3}S_n = S_{n+1}.\left.\right)$

2. Demuéstrese que $(1+x)^n \geq 1 + nx$, para $x > -1$.

3. Demuéstrese por inducción que $1 + 2 + \cdots + n = \frac{1}{2}n(n+1)$.

***4.** Demuéstrense por inducción las proposiciones siguientes:

(*a*) $$1 + 2q + 3q^2 + \cdots + nq^{n-1} = \frac{1-(n+1)q^n + nq^{n+1}}{(1-q)^2}.$$

(*b*) $$(1+q)(1+q^2)\cdots(1+q^{2^n}) = \frac{1-q^{2^{n+1}}}{1-q}.$$

5. Demuéstrese que, para todos los números naturales n mayores que 1, o bien n es un número primo o bien se puede expresar como producto de primos. *(Indicación:* sea A_{n-1} la afirmación para todo entero k tal que $k \leq n$, que k es primo o se puede expresar como producto de primos.)

***6.** Considérese la sucesión de fracciones

$$\frac{1}{1}, \frac{3}{2}, \frac{7}{5}, \ldots, \frac{p_n}{q_n}, \ldots,$$

donde $P_{n+1} = p_n + 2q_n$ y $q_{n+1} = p_n + q_n$.

(*a*) Demuéstrese para toda n que p_n/q_n está en sus menores términos.

(*b*) Demuéstrese que el valor absoluto de la diferencia entre p_n/q_n y $\sqrt{2}$ se puede hacer arbitrariamente pequeño. Demuéstrese también que el error de aproximación a $\sqrt{2}$ alterna de signo.

7. Sean a, b, a_n y b_n enteros tales que

$$(a + b\sqrt{2})^n = a_n + b_n\sqrt{2},$$

donde a es el entero más cercano a $b\sqrt{2}$. Demuéstrese que a_n es el entero que más se aproxima a $b_n\sqrt{2}$.

***8.** Definamos a_n y b_n por

$$a_1 = 3, \qquad a_{n+1} = 3^{a_n}, \qquad \text{y} \qquad b_1 = 9, \qquad b_{n+1} = 9^{b_n}.$$

Determínese para cada valor de n el valor mínimo m tal que $a_m \geq b_n$.

9. Si n es un número natural, demuéstrese que

$$\frac{(1+\sqrt{5})^n - (1-\sqrt{5})^n}{2^n\sqrt{5}}$$

es un número natural.

10. Determínese el número máximo de regiones en las que se puede dividir un plano por n rectas. Demuéstrese que el máximo ocurre cuando no hay dos rectas paralelas entre sí y no hay tres que concurran al mismo punto, y determínese el número de regiones que se tiene al permitir concurrencias y paralelismos.

11. Demuéstrese, para cada número natural n, que existe un número natural k tal que

$$(\sqrt{2} - 1)^n = \sqrt{k} - \sqrt{k-1}.$$

12. Demuéstrese inductivamente la desigualdad de Cauchy.

nSECCION 1.6, página 84

1. Demuéstrese que $\lim_{n\to\infty} (\sqrt{n+1} - \sqrt{n})(\sqrt{n+\tfrac{1}{2}}) = \tfrac{1}{2}$.

2. Demuéstrese que $\lim_{n\to\infty} (\sqrt[3]{n+1} - \sqrt[3]{n}) = 0$.

3. Sea $a_n = 10^n/n!$. (*a*) ¿A qué límite converge a_n? (*b*) ¿Es monótona la sucesión? (*c*) ¿Es monótona a partir de un cierto valor de n? (*d*) Dese una estimación de la diferencia entre a_n y el límite. (*e*) ¿A partir de qué valor de n es la diferencia menor que $1/100$?

4. Demuéstrese que $\lim_{n\to\infty} \dfrac{n!}{n^n} = 0$.

5. (*a*) Demuéstrese que $\lim_{n\to\infty} \left(\dfrac{1}{n^2} + \dfrac{2}{n^2} + \cdots + \dfrac{n}{n^2}\right) = \tfrac{1}{2}$.

(*b*) Demuéstrese que $\lim_{n\to\infty} \left(\dfrac{1}{n^2} + \dfrac{1}{(n+1)^2} + \cdots + \dfrac{1}{(2n)^2}\right) = 0$.

(*Indicación:* compárese esta suma con su término mayor.)

(*c*) Demuéstrese que $\lim_{n\to\infty} \left(\dfrac{1}{\sqrt{n}} + \dfrac{1}{\sqrt{n+1}} + \cdots + \dfrac{1}{\sqrt{2n}}\right) = \infty$.

*(*d*) Demuéstrese que $\lim_{n\to\infty} \left(\dfrac{1}{\sqrt{n^2+1}} + \dfrac{1}{\sqrt{n^2+2}} + \cdots + \dfrac{1}{\sqrt{n^2+n}}\right) = 1$.

6. Demuéstrese que todo decimal periódico representa un número racional. (Compárese con sección 1.1c, problema 3.)

7. Demuéstrese que $\lim_{n\to\infty} \dfrac{n^{100}}{1.01^n}$ existe y determínese su valor.

8. Demuéstrese que si a y $b \le a$ son positivos, la sucesión $\sqrt[n]{a^n + b^n}$ converge a a. De la misma manera, para cualesquiera k números positivos fijos a_1, $a_2, \ldots, a_k$ demuéstrese que $\sqrt[n]{a_1{}^n + a_2{}^n + \cdots + a_k{}^n}$ converge y encuéntrese su límite.

9. Demuéstrese que la sucesión $\sqrt{2}, \sqrt{2\sqrt{2}}, \sqrt{2\sqrt{2\sqrt{2}}}, \ldots$, converge. Encuéntrese el límite.

10. Si $\nu(n)$ es el número de factores primos que tiene n, demuéstrese que $\lim_{n\to\infty} \dfrac{\nu(n)}{n} = 0$.

11. Demuéstrese que si $\lim_{n\to\infty} a_n = \xi$, entonces $\lim_{n\to\infty} \sigma_n = \xi$, donde σ_n es la media aritmética $(a_1 + a_2 + \cdots + a_n)/n$.

12. Encuéntrense

(*a*) $\lim_{n\to\infty} \left(\dfrac{1}{1\cdot 2} + \dfrac{1}{2\cdot 3} + \cdots + \dfrac{1}{n(n+1)}\right)$

$\left(\text{\textit{Indicación:}}\ \dfrac{1}{k(k+1)} = \dfrac{1}{k} - \dfrac{1}{k+1}\right)$

(*b*) $\lim_{n\to\infty}\left(\frac{1}{1\cdot 2\cdot 3}+\frac{1}{2\cdot 3\cdot 4}+\cdots+\frac{1}{n(n+1)(n+2)}\right)$

13. Si $a_0 + a_1 + \cdots + a_p = 0$, demuéstrese que

$$\lim_{n\to\infty}(a_0\sqrt{n} + a_1\sqrt{n+1} + \cdots + a_n\sqrt{n+p}) = 0.$$

(Indicación: tómese $\sqrt{n}$ como factor común.)

14. Demuéstrese que $\lim_{n\to\infty} \sqrt[(2n+1)]{(n^2+n)} = 1$.

***15.** Sea a_n una sucesión tal que la sucesión $b_n = pa_n + qa_{n+1}$, donde $|p| < q$, es convergente. Probar que a_n converge. Si $|p| \geqslant q > 0$, mostrar que a_n no converge necesariamente.

16. Demuéstrese la validez de la relación

$$\lim_{n\to\infty}\frac{1}{n^{k+1}}\sum_{i=1}^{n} i^k = \frac{1}{k+1}$$

para cualquier entero k no negativo. *(Indicación:* aplíquese la inducción con respecto a k y empléese la relación

$$\sum_{i=1}^{n}[i^{k+1} - (i-1)^{k+1}] = n^{k+1},$$

con el desarrollo de $(i-1)^{k+1}$ en las potencias de i.)

SECCION 1.7, página 93

***1.** Sean a_1 y b_1 dos números positivos cualesquiera y sea $a_1 < b_1$. Definamos a_2 y b_2 por las ecuaciones

$$a_2 = \sqrt{a_1 b_1}, \qquad b_2 = \frac{a_1 + b_1}{2}.$$

De la misma manera, sean

$$a_3 = \sqrt{a_2 b_2}, \qquad b_3 = \frac{a_2 + b_2}{2},$$

y, en general,

$$a_n = \sqrt{a_{n-1}b_{n-1}}, \qquad b_n = \frac{a_{n-1} + b_{n-1}}{2}.$$

Demuéstrese (*a*) que la sucesión $a_1, a_2, \ldots$, converge, (*b*) que la sucesión $b_1, b_2, \ldots$, converge y (*c*) que las dos sucesiones tienen el mismo límite. (Este límite se llama *media aritmético-geométrica* de a_1 y b_1.)

***2.** Demuéstrese que el límite de la sucesión

$$\sqrt{2}, \sqrt{2+\sqrt{2}}, \sqrt{2+\sqrt{2+\sqrt{2}}}, \ldots$$

(*a*) existe y (*b*) es igual a 2.

***3.** Demuéstrese que el límite de la sucesión

$$a_n = \frac{1}{n} + \frac{1}{n+1} + \cdots + \frac{1}{2n}$$

existe. Dedúzcase que el límite es menor que 1 pero no menor que $\frac{1}{2}$.

4. Demuéstrese que el límite de la sucesión

$$b_n = \frac{1}{n+1} + \cdots + \frac{1}{2n}$$

existe y es igual al límite del ejemplo anterior.

5. Obténganse las cotas siguientes del límite L de los dos ejemplos anteriores: $37/60 < L < 57/60$.

***6.** Sean a_1, b_1 dos números positivos cualesquiera tales que $a_1 \leq b_1$. Sean

$$a_2 = \frac{2a_1 b_1}{a_1 + b_1}, \qquad b_2 = \sqrt{a_1 b_1},$$

y en general

$$a_n = \frac{2a_{n-1} b_{n-1}}{a_{n-1} + b_{n-1}}, \qquad b_n = \sqrt{a_{n-1} b_{n-1}}.$$

Demuéstrese que las sucesiones $a_1, a_2, \ldots$ y $b_1, b_2, \ldots$ convergen y tienen el mismo límite.

***7.** Demuéstrese que $1/e = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^n}{n!} + \cdots$. *(Indicación:* considérese el producto de las n-ésimas sumas parciales de los desarrollos de e y $1/e$.)

8. (a) Sin recurrir al teorema del binomio, demuéstrese que $a_n = (1 + 1/n)^n$ es monótona creciente y que $b_n = (1 + 1/n)^{n+1}$ es monótona decreciente. *(Indicación:* considérense a_{n+1}/a_n y b_n/b_{n+1}. Aplíquese el resultado de la sección 1.5, problema 2.)

(b) ¿Qué número es mayor $(1{,}000{,}000)^{1{,}000{,}000}$ ó $(1{,}000{,}001)^{999{,}999}$?

9. (a) A partir de los resultados del problema 8a demuéstrese que

$$\left(\frac{n}{e}\right)^n < n! < e(n+1)\left(\frac{n}{e}\right)^n.$$

(b) Para $n > 6$ dedúzcase la desigualdad más fuerte

$$n! < n\left(\frac{n}{e}\right)^n.$$

***10.** Si $a_n > 0$ y $\lim\limits_{n\to\infty} \frac{a_{n+1}}{a_n} = L$, entonces $\lim\limits_{n\to\infty} \sqrt[n]{a_n} = L$.

11. Por medio del problema 10, evalúense los límites de las sucesiones siguientes:

(a) $\sqrt[n]{n}$, (b) $\sqrt[n]{n^5 + n^4}$, (c) $\sqrt[n]{\frac{n!}{n^n}}$

12. Por medio del problema 11c, demuéstrese que

$$n! = n^n e^{-n} a_n,$$

donde a_n es un número cuya n-ésima raíz tiende a 1. (Véase el Apéndice, capítulo 7.)

13. (a) Evalúese

$$\frac{1}{1 \cdot 3} + \frac{1}{2 \cdot 4} + \cdots + \frac{1}{n(n+2)}.$$

(Indicación: compárese con la sección 1.6, problema 12a.)

(b) A partir del resultado anterior, demuéstrese que $\sum\limits_{k=1}^{\infty} \frac{1}{k^2}$ converge.

14. Sean p y q números naturales arbitrarios. Evalúense

(a) $\sum_{k=1}^{n} \frac{1}{(k+p)(k+p+q)}$.

(b) $\sum_{k=1}^{n} \frac{1}{(k+p)(k+p+q)}$.

15. Evalúense

(a) $\frac{1}{1\cdot 2\cdot 3} + \frac{1}{2\cdot 3\cdot 4} + \cdots + \frac{1}{n(n+1)(n+2)}$

(b) $\sum_{k=1}^{n} \frac{1}{k(k+1)(k+3)}$.

(c) Evalúense los límites de las expresiones anteriores cuando $n \to \infty$.

$*(d)$ Sean $a_1, a_2, \ldots, a_m$ enteros no negativos donde $a_1 < a_2 < \cdots < a_m$. Señálese la manera de obtener una fórmula de

$$S_n = \sum_{k=1}^{n} \frac{1}{(k+a_1)(k+a_2)\cdots(k+a_m)}$$

y de encontrar $\lim_{n\to\infty} S_n$.

16. Si a_k es monótona y $\sum_{k=1}^{\infty} a_k$ converge, demuéstrese que $\lim_{k\to\infty} ka_k = 0$.

17. Si a_k es monótona decreciente con límite 0 se tiene que $b_k = a_k - 2a_{k+1} + a_{k+2} \geq 0$ para toda k, demuéstrese entonces que $\sum_{k=1}^{\infty} kb_k = a_1$.

SECCION 1.8, página 105

1. Demuéstrese que $\lim_{m\to\infty} (\cos \pi x)^{2m}$ existe para cada valor de x y es igual a 1 o 0 según sea x entero o no lo sea.

2. (a) Demuéstrese que $\lim_{n\to\infty} [\lim_{m\to\infty} (\cos n!\,\pi x)^{2m}]$ existe para cada valor de x y es igual a 1 o 0 según sea x racional o irracional.

(b) Examínese la cuestión de la continuidad de estas funciones límite.

3. Sea $f(x)$ continua para $0 \leq x \leq 1$. Supongamos además que $f(x)$ no toma sino valores racionales y que $f(x) = \frac{1}{2}$ cuando $x = \frac{1}{2}$. Demuéstrese que $f(x) = \frac{1}{2}$ en todas partes.

SECCION 1.S.1, página 112

1. Sean $r = p/q$, $s = m/n$ números racionales arbitrarios, en los que p, q, m, n son enteros y q, n son positivos. Defínanse, en términos de los enteros p, q, m, n,

(a) $r + s$, (b) $r - s$, (c) rs, (d) $\frac{r}{s}$, (e) $r < s$.

2. Demuéstrese para encajes de intervalos racionales, $[a_n, b_n]$ y $[a_n', b_n']$, que cada una de las condiciones siguientes es necesaria y suficiente para tener la equivalencia:

(a) $a_n' - a_n$ es una sucesión nula,

(b) $a_n \leq b_n'$ y $a_n' \leq b_n$.

3. Dados $x \sim \{[a_n, b_n]\}$, $y \sim \{[\alpha_n, \beta_n]\}$, (a) verifíquese que las definiciones de adición y substracción,

$$x + y = \{[a_n + \alpha_n, b_n + \beta_n]\}, \qquad x - y = \{[a_n - \beta_n, b_n - \alpha_n]\},$$

tienen sentido. En particular, verifíquese que

(i) las representaciones dadas constituyen, de hecho, encajes de conjuntos para $x + y$ y $x - y$ cuando x e y son racionales;

(ii) si $x < y$, entonces $x + z < y + z$, donde z es un número real arbitrario.

(*b*) Defínase el producto xy y verifíquese sobre todo que la definición de producto tiene sentido, y que

(i) el conjunto encajado que se da constituye, en efecto, un encaje para xy cuando x e y son racionales.

(ii) si $x < y$ y $z > 0$, entonces $xz < yz$.

4. Demuéstrese que los principios siguientes son equivalentes en el sentido de que uno cualquiera de ellos se puede deducir como consecuencia de cualquier otro.

(*a*) Todo encaje de intervalos con extremos reales contiene un número real.

(*b*) Toda sucesión monótona y acotada converge.

(*c*) Toda sucesión infinita y acotada tiene por lo menos un punto de acumulación o límite.

(*d*) Toda sucesión de Cauchy converge.

(*e*) Todo conjunto acotado de números reales tiene un ínfimo y un supremo.

Problemas misceláneos

1. Si $w_1, w_2, \ldots, w_n > 0$, demuéstrese que el promedio ponderado

$$\frac{w_1x_1 + w_2x_2 + \cdots + w_nx_n}{w_1 + w_2 + \cdots w_2}$$

queda entre la mayor y la menor de las x.

2. Demuéstrese que

$$2(\sqrt{n} + 1 - 1) < 1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \cdots + \frac{1}{\sqrt{n}} < 2\sqrt{n}.$$

3. Demuéstrese para $x, y > 0$, que

$$\frac{x^n + y^n}{2} \geq \left(\frac{x + y}{2}\right)^n.$$

Interprétese geométricamente este resultado en términos de la gráfica de x^n.

4. Si $a_1 \geq a_2 \geq \cdots \geq a_n$ y $b_1 \geq b_2 \geq \cdots \geq b_n$, probar que

$$n\sum_1^n a_ib_i \geq \left(\sum_{i=1}^n a_i\right)\left(\sum_{i=1}^n b_i\right).$$

5. (*a*) Demuéstrese que la sucesión $a_1, a_2, a_3, \ldots$ se puede expresar como la sucesión de las sumas parciales de la serie $u_1, u_2, u_3, \ldots$ donde $u_n = a_n - a_{n-1}$ para $n \geq 1$ y $u_1 = a_1$.

(*b*) Exprésese la sucesión $a_n = n^3$ como sucesión de sumas parciales de una serie.

(*c*) Del resultado anterior, obténgase una fórmula de la suma parcial n-ésima de la serie

$$1 + 4 + 9 + \cdots + n^2 + \cdots.$$

(*d*) De la fórmula de $1^2 + 2^2 + \cdots + n^2$, obténgase una fórmula para

$$1^2 + 3^2 + 5^2 + \cdots + (2n + 1)^2.$$

6. Una sucesión se llama progresión aritmética de primer orden si las diferencias entre los términos sucesivos son constantes. Se llama progresión aritmética de

segundo orden si las diferencias de términos sucesivos constituyen una progresión aritmética de primer orden; y, en general, se llama progresión aritmética de orden k si las diferencias de términos sucesivos constituyen una progresión aritmética de orden $(k-1)$.

Los números 4, 6, 13, 27, 50, 84 son los primeros seis términos de una progresión aritmética. ¿Cuál es el menor orden posible? ¿Cuál es el octavo término de la progresión del menor orden con estos términos iniciales?

7. Demuéstrese que el n-ésimo término de una progresión aritmética de segundo orden se puede expresar en la forma $an^2 + bn + c$, donde a, b, c son independientes de n.

***8.** Demuéstrese que el n-ésimo término de una progresión aritmética de orden k se puede expresar en la forma $an^k + bn^{k-1} + \cdots + pn + q$, donde $a, b, \ldots, p$, q son independientes de n.

Encuéntrese el término n-ésimo de la progresión del menor orden en el problema 6.

9. Encuéntrese una fórmula del término n-ésimo de las progresiones aritméticas de menor orden de las que los términos iniciales son los siguientes:

(*a*) 1, 2, 4, 7, 11, 16, ...

(*b*) −7, −10, −9, 1, 25, 68, ...

***10.** Demuéstrese que la suma de los n primeros términos de una progresión aritmética de orden k es

$$a_kS_k + a_{k-1}S_{k-1} + \cdots + a_1S_1 + a_0n,$$

donde S_ν representa la suma de las n primeras potencias ν-ésimas y las a_i son independientes de n. Empléese este resultado para evaluar las sumas de las progresiones aritméticas del problema 9.

11. Al sumar

$$\nu(\nu+1)(\nu+2)\cdots(\nu+k+1) - (\nu-1)\nu(\nu+1)\cdots(\nu+k)$$

desde $\nu = 1$ hasta $\nu = n$, demuéstrese que

$$\sum_{\nu=1}^{n} \nu(\nu+1)(\nu+2)\cdots(\nu+k) = \frac{n(n+1)\cdots(n+k+1)}{k+2}.$$

12. Evalúese $1^3 + 2^3 + \cdots + n^3$ por medio de la relación

$$\nu^3 = \nu(\nu+1)(\nu+2) - 3\nu(\nu+1) + \nu.$$

13. Demuéstrese que la función

$$f(x) = \begin{cases} \dfrac{1}{\log_2 |x|}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

es continua, aunque sin la continuidad de Hölder. (*Indicación:* mostrar que la continuidad de Hölder con exponente α no se cumple en el origen, al considerar los valores $x - 1/2^{n/\alpha}$.)

14. Sea a_n una sucesión monótona decreciente de números no negativos. Demuéstrese que $\sum_{n=1}^{\infty} a_n$ converge si y sólo si lo hace $\sum_{\nu=0}^{\infty} 2^\nu a_{2^\nu}$.

15. En lo que sigue, investíguese la convergencia y determínese el límite cuando ello sea posible.

(*a*) $n!\,e - [n!\,e]$

(*b*) a_n/a_{n+1}, donde $a_1 = 0$, $a_2 = 1$, y $a_{k+2} = a_{k+1} + a_k$.

2

Las ideas fundamentales del cálculo integral y diferencial

Los procesos de límite fundamentales del cálculo son la integración y la derivación o diferenciación. Casos aislados de estos procesos del cálculo (que culminaron con el trabajo de Arquímedes) fueron considerados aún en la antigüedad, y con frecuencia creciente en los siglos XVI y XVII. Sin embargo, el desarrollo sistemático del cálculo, no iniciado sino hasta en el siglo XVII, es usualmente atribuido a los dos grandes pioneros de la ciencia, Newton y Leibnitz. La clave para este desarrollo sistemático es la comprensión de que los procesos de derivación e integración, que fueron tratados separadamente, están íntimamente relacionados al ser recíprocos uno del otro.[1]

Una evaluación histórica imparcial de los méritos no puede atribuir la invención del cálculo a repentinos destellos inexplicables de genio por parte de uno o dos individuos. Muchas gentes, tales como Fermat, Galileo y Kepler, estimulados por las nuevas ideas revolucionarias en la ciencia, contribuyeron a los fundamentos del cálculo. De hecho, el profesor de Newton, Barrow, estuvo casi en posesión completa de la comprensión básica de la reciprocidad entre derivación e integración, que es la piedra angular del cálculo sistemático de Newton y Leibnitz. Newton expuso los conceptos de manera un poco más clara; por otra parte, la ingeniosa notación de Leibnitz y sus métodos de cálculo son altamente sugestivos y siguen siendo indispensables. El trabajo de estos dos hombres estimuló inmediatamente las ramas superiores del análisis, incluyendo el cálculo de variaciones y la teoría de ecuaciones diferenciales, y condujo a innumerables aplicaciones en las ciencias naturales. Muy curioso resulta que, no obstante que Newton, Leibnitz y sus sucesores inmediatos hicieron tales variados usos de la po-

[1] Este enunciado constituye el "teorema fundamental del cálculo".

derosa herramienta puesta en sus manos, ninguno consiguió clarificar completamente los conceptos básicos involucrados en sus trabajos. Sus argumentos emplearon "cantidades infinitamente pequeñas" en formas que son lógicamente insostenibles y no convincentes. La clarificación aconteció a fines del siglo XIX con la cuidadosa formulación del concepto de límite y con el análisis del continuo de números, como se explica en el capítulo 1.[1]

Comenzaremos con una discusión de los conceptos fundamentales. Estos pueden apreciarse totalmente sólo a través de ilustraciones concretas y ejemplos; es, por lo tanto, que se recomienda aquí, como en muchas partes en este libro, que las secciones teóricas y generales sean cuidadosamente estudiadas nuevamente después de que el lector haya absorbido el material más específico y concreto en secciones subsiguientes.

2.1 La integral

a. Introducción

Sólo después de un prolongado desarrollo los procedimientos sistemáticos de integración y derivación alcanzaron la necesidad de descripciones matemáticas precisas para nociones intuitivas que surgen en la geometría y en las ciencias naturales. La derivación es el concepto requerido para describir las nociones de *tangentes* a curvas y el de velocidad de partículas en movimiento, o bien, más generalmente, el concepto de *razón de variación o cambio*. El concepto intuitivo de *área* de una región con fronteras curvas encuentra su formulación matemática precisa en el proceso de integración. Muchos otros conceptos relacionados en la geometría y en la física requieren también de la integración, como se verá posteriormente. En esta sección se introduce el concepto de integral en conexión con el problema de medir el área de una región plana acotada por curvas.

Areas. Se tiene una percepción intuitiva de que una región contenida dentro de una curva cerrada posee un "área" la cual mide el número de unidades cuadradas dentro de la curva. Empero, la pregunta de *cómo* esta medida para el área puede ser descrita en términos precisos exige una cadena de pasos matemáticos. Las propiedades básicas del área que la intuición sugiere son: el área es un número (positivo, dependiente de la elección de la unidad de longitud); este número es el mismo para figuras congruentes; para todos los rectángulos es el producto de las longitudes de

---

[1] El surgimiento del cálculo, que se prolonga por más de 2,000 años, representa uno de los capítulos más fascinantes en la historia del descubrimiento científico. A los lectores interesados se les recomienda el libro de Carl B. Boyer, *Concept of the Calculus*, Hafner Publishing Company, 1949. Véase también O. Toeplitz, *Calculus*, A Genetic Approach, University of Chicago, 1963.

los lados adyacentes; y, finalmente, para una región descompuesta en secciones el área del todo es igual a la suma de las áreas de las secciones.

Una consecuencia inmediata es el hecho de que: para una región A que es parte de una región B, el área de A no puede exceder el área de B.

Estas propiedades permiten el cálculo directo del área de cualquier figura que puede ser descompuesta en un número finito de rectángulos. Más generalmente, para asignar un valor F al área de una región R, consideramos otras dos regiones R' (inscripta) y R'' (circunscripta) que se puedan descomponer en rectángulos, donde R'' contiene a R y R' está contenida en R

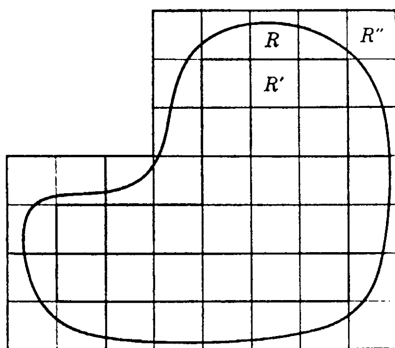


Figura 2.1 Aproximación de un área.

(ver Fig. 2.1). Se sabe entonces, por lo menos, que F debe encontrarse entre las áreas de R' y R''. El valor de F está completamente determinado si se encuentran sucesiones de regiones circunscriptas R_n'' y regiones inscriptas R_n' que puedan ambas descomponerse en rectángulos y tales que las áreas de R_n'' y R_n' tengan el mismo límite cuando n tiende a infinito. Este es, remontándonos a la antigüedad, el método de "exhaucción", el cual es usado en geometría elemental para describir el área de un círculo.[1] La formulación precisa de esta idea intuitiva conduce ahora a la noción de integración.

b. La integral como un área

Area bajo una curva

La noción analítica de integral surge cuando se asocian áreas con *funciones:* Considérese el área de una región acotada por la izquierda y por

[1] Por supuesto, puede utilizarse cualquier clase de *polígonos* inscriptos y circunscritos, puesto que un polígono puede ser descompuesto en triángulos rectángulos y el área de un triángulo rectángulo es claramente la mitad de la de un rectángulo con los mismos lados.

la derecha mediante líneas verticales $x = a$ y $x = b$, debajo por el eje x y arriba por la gráfica de una función positiva continua $f(x)$ (Fig. 2.2). En breve, nos referiremos a esto como el área "bajo la curva". Por el mbmento se acepta como intuitiva la idea de que el área de tal región es un número definido. Llamamos esta área F_a^b la *integral de la función* f entre los lí-

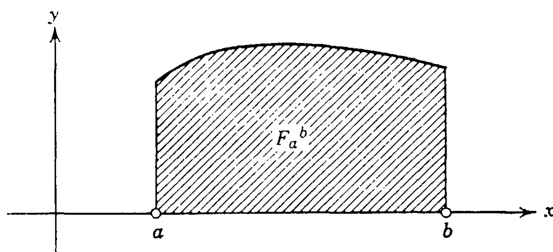


Figura 2.2.

mites[1] a y b. En la investigación del valor numérico de F_a^b se hace uso de aproximaciones mediante sumas de áreas de rectángulos. Para este propósito se divide el intervalo (a, b) del eje x en n partes (pequeñas), no necesariamente del mismo tamaño, las cuales se denominarán *celdas*. En cada punto de división se traza la línea perpendicular al eje x, hacia arriba hasta la curva. La región con área F_a^b es así dividida en n franjas, cada una

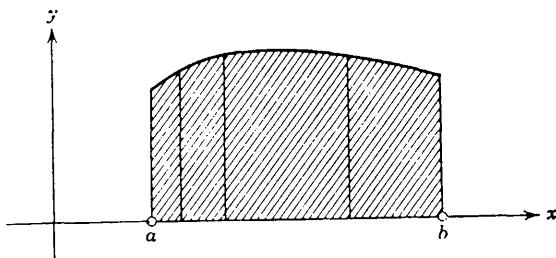


Figura 2.3.

acotada por una porción de la gráfica de la función $f(x)$ y por tres segmentos de recta (Fig. 2.3).

Area o integral como límite de una suma. El cálculo preciso del área de tales franjas no es más fácil que el cálculo del área de la región original. Es, sin embargo, dar un paso hacia adelante el aproximar el área de cada franja desde arriba y desde abajo mediante las áreas de los rectángulos circunscritos e inscritos con la misma base, donde la frontera curva de la

---

[1] No debe surgir confusión del uso de la palabra "límite" para puntos frontera del intervalo de integración.

franja se reemplaza por una línea horizontal a una distancia del eje x que es o bien el máximo o bien el mínimo valor de $f(x)$ en la celda (Fig. 2.4). Más generalmente, se obtiene una aproximación intermedia si se reemplaza la franja por un rectángulo de la misma base y acotado en la parte superior por una línea horizontal que intersecta la frontera curva de la franja (véase Fig. 2.5). Analíticamente, esto equivale a reemplazar la fun-

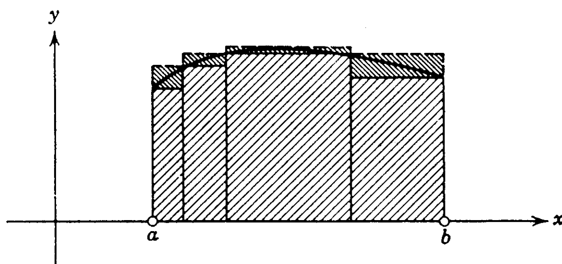


Figura 2.4.

ción $f(x)$ en cada una de las celdas por algún valor intermedio constante. Denótese por F_n la suma de las n áreas rectangualres. La intuición nos dice que los valores F_n tienden a $F_a{}^b$ si la subdivisión se hace cada vez más fina, esto es, si se hace crecer n sin límite mientras que la máxima longitud de las celdas individuales tiende a cero. De esta forma $F_a{}^b$ está representado como un límite de áreas que consisten de rectángulos.

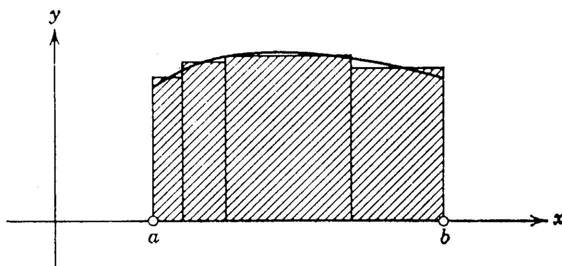


Figura 2.5.

c. Definición analítica de la integral. Notaciones

Definición y existencia de integrales

En el párrafo anterior aceptamos al área bajo una curva como una cantidad dada intuitivamente, y a continuación ésta se representó como un valor límite. Ahora se invertirá el procedimiento. No se invocará más a la intuición para asignar un área a la región bajo una curva continua; por

el contrario, se comenzará de manera puramente analítica con las sumas F_n definidas anteriormente y se *probará* que estas sumas tienden a un límite definido. Este límite es entonces la *definición* precisa de la integral y del área.

Sea la función $f(x)$ continua (pero no necesariamente positiva) en el intervalo cerrado $a \leq x \leq b$. Se divide el intervalo por $(n-1)$ puntos $x_1, x_2, \ldots, x_{n-1}$ en n celdas iguales o distintas, con las longitudes

$$x_i - x_{i-1} = \Delta x_i, \qquad (i = 1, 2, \ldots, n)^1,$$

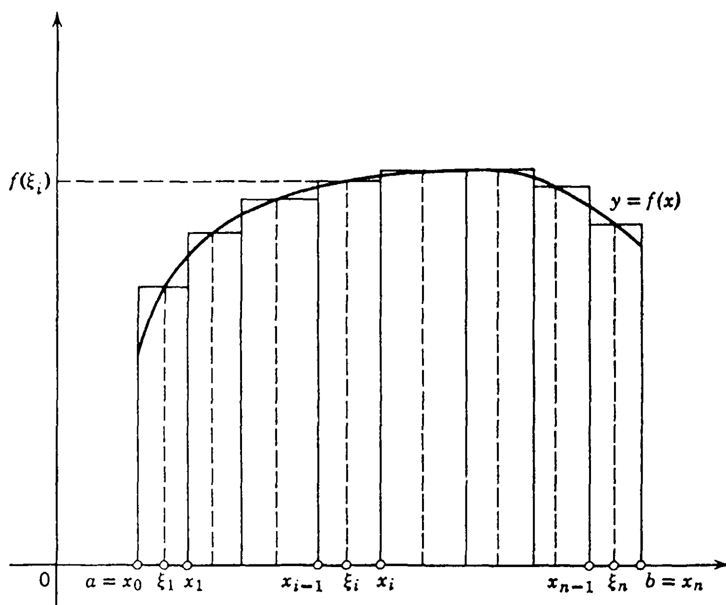


Figura 2.6 Para ilustrar la definición analítica de integral.

donde, además, se hace $x_0 = a$, $x_n = b$ (ver Fig. 2.6). En cada subintervalo cerrado $[x_{i-1}, x_i]$ o celda se elige cualquier punto ξ_i. Se forma la suma

$$\begin{aligned} F_n &= f(\xi_1)(x_1 - x_0) + f(\xi_2)(x_2 - x_1) + \cdots + f(\xi_n)(x_n - x_{n-1}) \\ &= f(\xi_1)\Delta x_1 + f(\xi_2)\Delta x_2 + \cdots + f(\xi_n)\Delta x_n. \end{aligned}$$

Utilizando el símbolo de sumatoria se escribe más concisamente

$$F_n = \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1})$$

[1] El símbolo Δ no debe ser interpretado como un factor sino solamente como indicador de una diferencia en valores de la variable que le sigue. Así, el símbolo Δx_i significa la diferencia $x_i - x_{i-1}$ de valores consecutivos de x.

o bien

$$F_n = \sum_{i=1}^{n} f(\xi_i)\Delta x_i.$$

Si $f(x)$ es positivo, el valor F_n representa el área bajo la curva, obtenida substituyendo f en cada subintervalo por el valor constante $f(\xi_i)$. Por supuesto, las sumas F_n pueden ser formadas sin suponer que f es positivo. Parece plausible intuitivamente que las sumas F_n deben tender a un límite F_a^b conforme el número n de intervalos crece indefinidamente y al mismo tiempo la longitud del máximo subintervalo tiende a cero. Esto implicaría que el valor del límite F_a^b es independiente de la manera particular según la cual los puntos de subdivisión $x_1, x_2, \ldots, x_{n-1}$ y los puntos intermedios $\xi_1, \xi_2, \ldots, \xi_n$ son elegidos. Denominamos a F_a^b la integral de $f(x)$ entre los límites a y b.

La intuición geométrica, no importa cuán convincente, puede servir solamente como guía a los procesos de límite analíticos; por lo tanto, una justificación analítica es necesaria y se debe proporcionar una demostración para la existencia de la integral como el límite descripto arriba. Además, como se ha dicho ya, no debe insistirse en la hipótesis de que la función f es positiva en el intervalo.

Así se afirma:

Teorema de Existencia. *Para toda función continua $f(x)$ en un intervalo cerrado $[a, b]$ la integral sobre este intervalo existe como el límite de las sumas F_n descritas arriba (independientemente de la elección de los puntos de subdivisión $x_1, \ldots, x_{n-1}$ y de los puntos intermedios $\xi_1, \ldots, \xi_n$, siempre que el máximo de las longitudes Δx_i tienda a cero).*

Se adquirirá primero cierta experiencia y comprensión antes de considerar la demostración de la existencia de la integral, lo que se hará en el Suplemento (p. 213).

Notación de Leibnitz para la integral

La definición de la integral como el límite de una suma condujo a Leibnitz a expresar la integral mediante el siguiente símbolo:

$$\int_a^b f(x)\, dx.$$

El signo de integral es una modificación del signo de sumatoria en forma de una S grande que se usó en la época de Leibnitz. El paso al límite a partir de una subdivisión finita en porciones Δx_i es indicada mediante el uso de la letra d en lugar de Δ. Sin embargo, al utilizarse esta notación no debe tolerarse el misticismo del siglo XVIII de considerar dx como un

"infinitamente pequeño" o "cantidad infinitesimal", o de considerar la integral como una "suma de un número infinito de cantidades infinitamente pequeñas". Tal concepción está desprovista de significado claro y oscurece lo que anteriormente se ha formulado con precisión. Desde el punto de vista presente, el símbolo individual dx no ha sido definido. La sugestiva combinación de símbolos $\int_a^b f(x)\,dx$ está definida para una función $f(x)$ en el intervalo $[a, b]$ mediante la formación de las sumas ordinarias F_n y pasando al límite con $n \to \infty$.

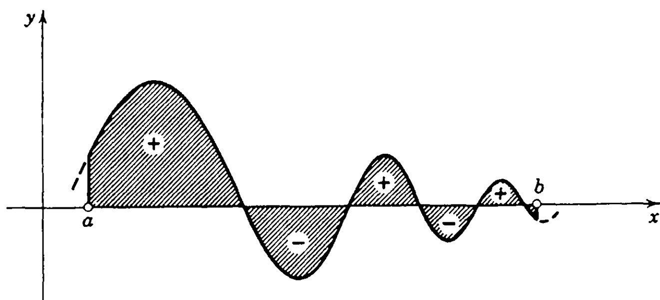


Figura 2.7.

El símbolo particular que se utiliza para la variable de integración es una cuestión sin importancia alguna (así como en la notación para sumas no importa cómo se denominó al índice de la sumatoria); en lugar de $\int_a^b f(x)\,dx$ puede igualmente escribirse $\int_a^b f(t)\,dt$ o bien $\int_a^b f(u)\,du$. El *integrando* denotado por f es una función de una variable independiente sobre el intervalo $[a, b]$ y el nombre de la variable es indiferente. Solamente los puntos extremos del intervalo de integración a y b afectan el valor de la integral para una f dada. Expresiones como $\int_a^b f(x)\,dx$ o bien $\int_a^b f(a)\,da$, en las cuales se utiliza la misma letra para la variable de integración y un punto extremo, son engañosas bajo nuestra definición y deberán, al principio, evitarse.

Si el *integrando* $f(x)$ es positivo en el intervalo $[a, b]$, puede inmediatamente identificarse $\int_a^b f(x)\,dx$ con el área acotada por la gráfica de f y las rectas $x = a$, $x = b$ e $y = 0$. Sin embargo, la integral de f está definida analíticamente como el límite de las sumas F_n independientemente de cualquier suposición acerca del signo de f. Si $f(x)$ es negativa en todo o parte del intervalo, el único efecto es el de hacer los correspondientes factores

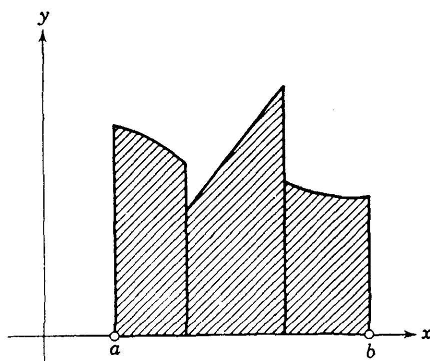


Figura 2.8.

$f(\xi_i)$ en la suma negativos en vez de positivos. A la región acotada por la parte de la curva debajo del eje x se asignará entonces naturalmente un área negativa. La integral será así la suma de términos positivos y negativos, correspondiendo respectivamente a porciones de la curva encima y debajo del eje x[1] (véase Fig. 2.7).

Es intuitivamente convincente afirmar que nuestro proceso de límite converge, aún si la función $f(x)$ no es continua en todas partes sino que

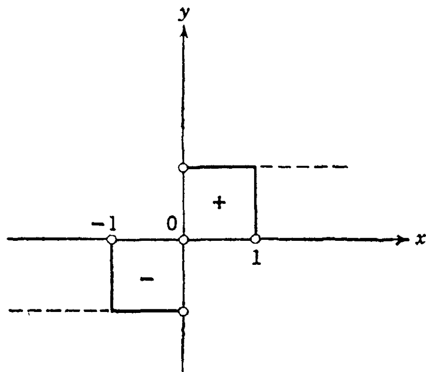


Figura 2.9 $\int_{-1}^{1} \operatorname{sgn} x \, dx = 0.$

posee discontinuidades de salto en uno o varios puntos, como la función indicada por la curva en la Fig. 2.8, donde claramente existe un área bajo la curva.[2]

---

[1] Las áreas de regiones acotadas por curvas cerradas arbitrarias se considerarán en el capítulo 4.

[2] Como otro ejemplo considérese $f(x) = \operatorname{sgn} x$ en $[-1, 1]$. Se tiene $f(x) = -1$ para $x < 0$ y $f(x) = +1$ para $x > 0$ (véase Fig. 2.9). Entonces $\int_{-1}^{+1} f(x)\,dx = 0$.

Así, el proceso de límite anterior puede muy bien resultar en un límite definido de la suma F_n para funciones que poseen algunas discontinuidades; y esta posibilidad se indica denominando a tales funciones *integrables*. A mediados del siglo XIX el gran Bernhard Riemann analizó por primera vez la aplicabilidad del proceso de integración a funciones generales. Más recientemente se han introducido varias extensiones del concepto mismo de integración. Empero, tales refinamientos poseen una menor importancia inmediata para el cálculo relativo a fenómenos intuitivamente accesibles, y no será necesario para nosotros destacar siempre la integrabilidad de nuestras funciones como mero recordatorio de que pueden definirse funciones no integrables.

En cursos avanzados la integral que se ha definido aquí se denomina "integral de Riemann", para distinguirla de varios conceptos generalizados de integral; las sumas de aproximación F_n se denominan sumas de *Riemann*.

2.2 Ejemplos elementales de integración

En un buen número de casos significativos estamos ya en condiciones de calcular la integral de una función llevando a cabo los procesos de límite prescriptos. Esto se hará mediante una evaluación explícita de las sumas F_n para una elección adecuada de los puntos intermedios ξ_i (usualmente, el punto extremo izquierdo o derecho de las celdas). El teorema sobre la existencia de la integral de una función continua asegura que el límite de los F_n es el mismo para cualquier otra elección de los puntos intermedios ξ_i y para cualquier método de subdivisión.

a. Integración de una función lineal

Primero se verificará que la integral realmente da el valor correcto del área para algunas figuras sencillas que conocemos de la geometría.

Sea $f(x)$ = constante = γ. Para calcular la integral de $f(x)$ entre los límites a y b se forman las sumas F_n (véase Fig. 2.10). Puesto que aquí $f(\xi_i) = \gamma$, se encuentra

$$F_n = \sum_{i=1}^{n} \gamma\, \Delta x_i = \gamma \sum_{i=1}^{n} \Delta x_i = \gamma(b-a).$$

Por lo tanto, de manera análoga

$$\lim_{n\to\infty} F_n = \int_a^b \gamma\, dx = \gamma(b-a).$$

Esta es justamente la fórmula para el área de un rectángulo de altura γ y base $b - a$.

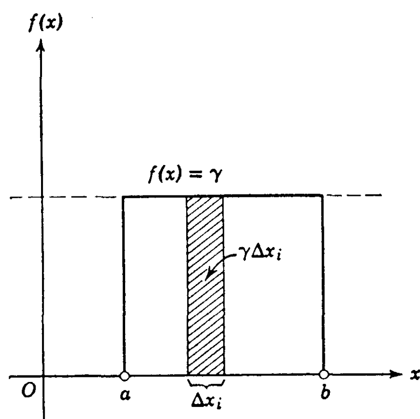


Figura 2.10 Integral de una constante.

La integral de la función $f(x) = x$,

$$\int_a^b x\, dx,$$

(Fig. 2.11), como se sabe de la geometría elemental, posee el valor

$$\tfrac{1}{2}(b - a)(b + a) = \tfrac{1}{2}(b^2 - a^2).$$

Para confirmar que el proceso límite conduce analíticamente al mismo resultado, se subdivide el intervalo de a a b en n partes iguales por medio de los puntos de división

$$a + h, a + 2h, \ldots, a + (n - 1)h,$$

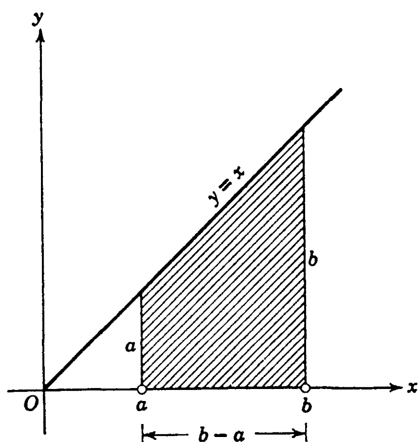


Figura 2.11.

donde $h = (b - a)/n$. Tomando para ξ_i el punto extremo derecho de cada intervalo se encuentra la integral como el límite, para $n \to \infty$, de la suma

$$F_n = (a + h)h + (a + 2h)h + \cdots + (a + nh)h$$
$$= nah + (1 + 2 + 3 + \cdots + n)h^2 = nah + \tfrac{1}{2}n(n + 1)h^2,$$

donde se ha utilizado la bien conocida fórmula para la suma de una progresión aritmética (véase p. 134, problema 3). Substituyendo $h = (b - a)/n$, se ve que

$$F_n = a(b - a) + \frac{1}{2}\left(1 + \frac{1}{n}\right)(b - a)^2,$$

de lo que se sigue inmediatamente que

$$\lim_{n\to\infty} F_n = a(b - a) + \tfrac{1}{2}(b - a)^2 = \tfrac{1}{2}(b^2 - a^2).$$

b. Integración de x^2

La geometría elemental no conduce tan fácilmente a la integración de la función $f(x) = x^2$, esto es, a la determinación del área de la región [1] acotada por un segmento de una parábola, un segmento del eje x y dos coordenadas. Es necesario un verdadero proceso de límite. Suponiendo $a < b$, se eligen los mismos puntos de división y los mismos puntos intermedios que en el ejemplo anterior (véase Fig. 2.12). Se sigue entonces que la integral de x^2 entre los límites a y b es el límite de las sumas

$$F_n = (a + h)^2h + (a + 2h)^2h + \cdots + (a + nh)^2h$$
$$= na^2h + 2ah^2(1 + 2 + 3 + \cdots + n)$$
$$+ h^3(1^2 + 2^2 + 3^2 + \cdots + n^2);$$

usando los valores conocidos de las sumas encerradas entre paréntesis se encuentra (véase pp. 81-82)

$$F_n = na^2h + n(n + 1)ah^2 + \frac{1}{6}[n(n + 1)(2n + 1)]h^3$$
$$= a^2(b - a) + \left(1 + \frac{1}{n}\right)a(b - a)^2 + \frac{1}{6}\left(1 + \frac{1}{n}\right)\left(2 + \frac{1}{n}\right)(b - a)^3.$$

Puesto que $\lim\limits_{n\to\infty} \frac{1}{n} = 0$, se tiene

$$\lim_{n\to\infty} F_n = a^2(b - a) + a(b - a)^2 + \frac{1}{3}(b - a)^3 = \frac{1}{3}(b^3 - a^3).$$

---

[1] Algunas veces se expresa esto diciendo que se está "cuadrando" la región.

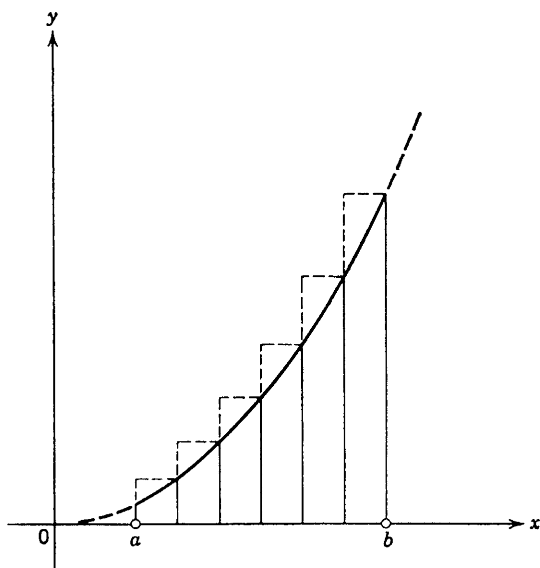


Figura 2.12 Area bajo un arco parabólico mediante subdivisión aritmética.

Así, para $a < b$,

$$\int_a^b x^2\, dx = \frac{1}{3}(b^3 - a^3).$$

*c. Integración de x^α para enteros $\alpha \neq -1$

Los siguientes ejemplos de esta sección son ilustraciones instructivas que muestran que en algunos casos la integración puede llevarse a cabo mediante recursos especiales elementales. Posteriormente, en la Sección 2.9d (p. 212) se obtendrán los mismos resultados de manera más simple utilizando métodos generales.

El mismo tipo de argumento que el utilizado para x y x^2, aplicado a las funciones x^3, x^4, . . . , da como resultado la relación

$$(1) \qquad \int_a^b x^\alpha\, dx = \frac{1}{\alpha + 1}(b^{\alpha+1} - a^{\alpha+1}),$$

donde α es cualquier entero positivo; esto puede ser demostrado encontrando fórmulas apropiadas para las sumas $1^\alpha + 2^\alpha + \cdots + n^\alpha$, tal como la relación

$$\lim_{n\to\infty}\left[(1^\alpha + 2^\alpha + \cdots + n^\alpha)\frac{1}{n^{\alpha+1}}\right] = \frac{1}{\alpha + 1}$$

que puede ser probada por inducción sobre α (véase el Problema 16, p. 136). En la siguiente sección la fórmula (1) será probada de una forma

diferente, con mayor generalidad y simplicidad, indicándose con ello la potencia de los métodos que desarrollaremos. Su validez será extendida a todos los valores de α excepto $\alpha = -1$.

Afortunadamente, la definición de integral nos deja bastante latitud en la elección de las subdivisiones y proporciona un camino más sencillo para evaluar la integral. No se tienen que utilizar sumas basadas en puntos equidistantes de división. En vez de eso, con el "cociente" $\sqrt[n]{b/a} = q$ se

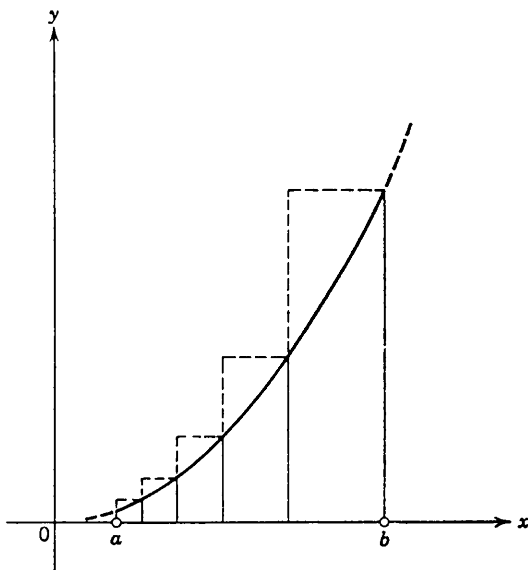


Figura 2.13 Area bajo un arco parabólico mediante subdivisión geométrica.

subdivide el intervalo $[a, b]$ mediante los puntos de una progresión geométrica (Fig. 2.13),

$$a, aq, aq^2, \ldots, aq^{n-1}, aq^n = b;$$

se requiere entonces solamente evaluar la suma de una serie geométrica. Dados los puntos de división $x_i = aq^i$ la longitud de la i-ésima celda está dada por

$$\Delta x_i = aq^i - aq^{i-1} = \frac{aq^i(q-1)}{q}.$$

El máximo Δx_i es el último:

$$\Delta x_n = \frac{b(q-1)}{q}.$$

Para $n \to \infty$ el número q tiende hacia el valor uno (véase Ejemplo *d*, p. 87), y por lo tanto la longitud Δx_n de la máxima celda, y entonces también

la longitud de todas las celdas, tiende a cero. Para los puntos intermedios ξ_i se eligen nuevamente los puntos extremos derechos x_i de cada celda. La suma

$$(2) \qquad F_n = \sum_{i=1}^{n} (\xi_i)^\alpha \Delta x_i = \sum_{i=1}^{n} (aq^i)^\alpha aq^i \frac{q-1}{q}$$

$$= a^{\alpha+1} \frac{q-1}{q} \sum_{i=1}^{n} (q^{1+\alpha})^i$$

es conocida explícitamente a partir de la suma de la progresión geométrica con razón $q^{1+\alpha}$. Aplicando la bien conocida fórmula (p. 91), se encuentra

$$F_n = a^{\alpha+1} \frac{q-1}{q} q^{\alpha+1} \frac{q^{n(\alpha+1)} - 1}{q^{\alpha+1} - 1}$$

$$= a^{\alpha+1}(q-1) q^\alpha \frac{(b/a)^{\alpha+1} - 1}{q^{\alpha+1} - 1} = (b^{\alpha+1} - a^{\alpha+1}) q^\alpha \frac{q-1}{q^{\alpha+1} - 1}.$$

Puesto que $q \neq 1$, puede utilizarse una vez más la fórmula para la suma de una progresión geométrica y escribir

$$\frac{q-1}{q^{1+\alpha} - 1} = \frac{1}{q^\alpha + q^{\alpha-1} + \cdots + 1}.$$

Para $n \to \infty$ todas las potencias de q tienden a uno y se sigue que

$$\lim_{n\to\infty} F_n = \frac{1}{1+\alpha} (b^{1+\alpha} - a^{1+\alpha}).$$

De esta forma se ha verificado la fórmula (1) para la integral de x^α para $0 < a < b$ y cualquier entero positivo α.

El mismo método se aplica también para enteros negativos α, siempre que $\alpha \neq -1$. Para la suma F_n se obtiene como antes

$$F_n = (b^{\alpha+1} - a^{\alpha+1}) q^\alpha \frac{q-1}{q^{\alpha+1} - 1}$$

$$= (b^{\alpha+1} - a^{\alpha+1}) \frac{q-1}{q(1 - q^{-\alpha-1})},$$

donde se recuerda que $-\alpha$ es positivo y mayor que uno. Aplicando la fórmula para una progresión geométrica, se obtiene

$$\frac{1}{q} \left(\frac{q-1}{q^{-\alpha-1} - 1} \right) = \frac{1}{q^{-\alpha-1} + q^{-\alpha-2} + \cdots + q}$$

que tiende a $1/(-\alpha - 1)$ conforme $n \to \infty$. Consecuentemente, como antes,

$$\lim_{n\to\infty} F_n = \frac{1}{\alpha+1} (b^{\alpha+1} - a^{\alpha+1}),$$

La fórmula integral no tiene significado para $\alpha = -1$, puesto que tanto el numerador como el denominador en el miembro derecho serían entonces cero. En vez de esto se encuentra, de la expresión original (2) para F_n para el caso $\alpha = -1$, que $F_n = n(q-1)/q$. Consecuentemente, observando que $q = \sqrt[n]{b/a}$ tiende a uno conforme $n \to \infty$, se encuentra

$$(3) \qquad \int_a^b \frac{1}{x}\,dx = \lim_{n\to\infty} n(\sqrt[n]{b/a} - 1).$$

Aquí el límite en el miembro derecho no puede ser expresado en términos de potencias de a y de b pero puede ser expresado en términos de logaritmos de esas cantidades como se verá posteriormente (p. 167).

*d. Integración de x^α para α racional distinto de -1

El resultado obtenido anteriormente puede ser considerablemente generalizado sin complicar esencialmente la demostración. Sea $\alpha = r/s$ un número racional positivo, siendo r y s enteros positivos; entonces en la evaluación de la integral dada arriba nada se cambia excepto la evaluación del límite $(q-1)/(q^{\alpha+1}-1)$ cuando q se aproxima a uno. Esta expresión es ahora simplemente $(q-1)/(q^{(r+s)/s}-1)$. Hagamos $q^{1/s} = \tau (\tau \neq 1)$: Entonces, si q tiende a uno τ también tiende a uno. Tenemos por ello que encontrar el valor límite de $(\tau^s - 1)/(\tau^{r+s}-1)$ cuando τ se aproxima a uno. Si se dividen tanto el numerador como el denominador por $\tau - 1$ y se transforman como antes mediante la fórmula para progresiones geométricas, el límite se torna simplemente

$$\lim_{\tau\to 1} \frac{\tau^{s-1} + \tau^{s-2} + \cdots + 1}{\tau^{r+s-1} + \tau^{r+s-2} + \cdots + 1}.$$

Puesto que tanto el numerador como el denominador son continuos en τ, este límite se obtiene de inmediato substituyendo $\tau = 1$, y por lo tanto se hace igual a $s/(r+s) = 1/(\alpha+1)$; luego, para todo valor racional positivo de α se obtiene la fórmula integral

$$\int_a^b x^\alpha\,dx = \frac{1}{\alpha+1}(b^{\alpha+1} - a^{\alpha+1}),$$

tal como con enteros positivos.

Esta fórmula permanece válida para valores negativos racionales de $\alpha = -r/s$, siempre que se excluya el valor $\alpha = -1$ (para el cual la fórmula utilizada arriba para la suma de la progresión geométrica pierde su significado).

Para α negativo se evalúa nuevamente el límite de $(q-1)/(q^{\alpha+1}-1)$ haciendo $q^{-1/s} = \tau$ para $\alpha = -r/s$; esto se deja como un ejercicio para el lector.

Es natural conjeturar que el campo de validez de la última fórmula se extiende también a valores irracionales de α. En realidad, se establecerá la fórmula integral para todos los valores reales de α (excepto $\alpha = -1$) en la sección 2.7 (p. 176) de una manera enteramente simple como una consecuencia de la teoría general.

*e. Integración de sen x y cos x

El último ejemplo elemental que será tratado aquí mediante un artificio especial es la integral de $f(x) = \operatorname{sen} x$. La integral

$$\int_a^b \operatorname{sen} x \, dx$$

es claramente el límite de la suma

$$S_h = h[\operatorname{sen}(a + h) + \operatorname{sen}(a + 2h) + \cdots + \operatorname{sen}(a + nh)],$$

que surge de la división del intervalo de integración en celdas de longitud $h = (b - a)/n$. Multiplíquese la expresión en el miembro derecho por $2 \operatorname{sen} h/2$ y recuérdese la bien conocida fórmula trigonométrica

$$2 \operatorname{sen} u \operatorname{sen} v = \cos(u - v) - \cos(u + v).$$

Siempre que h no sea un múltiplo de 2π se obtiene la fórmula

$$S_h = \frac{h}{2\operatorname{sen}\frac{h}{2}}\left[\cos\left(a + \frac{h}{2}\right) - \cos\left(a + \frac{3}{2}h\right) + \cos\left(a + \frac{3}{2}h\right)\right.$$
$$\left. - \cos\left(a + \frac{5}{2}h\right) + \cdots + \cos\left(a + \frac{2n-1}{2}h\right) - \cos\left(a + \frac{2n+1}{2}h\right)\right]$$
$$= \frac{h}{2\operatorname{sen}\frac{h}{2}}\left[\cos\left(a + \frac{h}{2}\right) - \cos\left(a + \frac{2n+1}{2}h\right)\right].$$

Puesto que $a + nh = b$, la integral se convierte en el límite de

$$\frac{h}{2\operatorname{sen}\frac{h}{2}}\left[\cos\left(a + \frac{h}{2}\right) - \cos\left(b + \frac{h}{2}\right)\right] \quad \text{conforme} \quad h \to 0.$$

Se sabe ahora, del capítulo 1 (p. 107), que para $h \to 0$ la expresión $(h/2)/(\operatorname{sen} h/2)$ se aproxima al límite uno. El límite deseado es entonces simplemente $\cos a - \cos b$, y se obtiene la integral

$$\int_a^b \operatorname{sen} x \, dx = -(\cos b - \cos a).$$

Análogamente,

$$\int_a^b \cos x \, dx = \operatorname{sen} b - \operatorname{sen} a \qquad \text{(véase problema 3, p. 217).}$$

Cada uno de los ejemplos anteriores fue tratado con un artificio especial. Empero, el punto esencial del cálculo integral y diferencial sistemáticos es el simple hecho de que en vez de tales artificios especiales se usan consideraciones generales que conducen directamente al resultado. Llegaremos a estos métodos discutiendo primero algunas reglas generales concernientes a integrales, introduciendo luego el concepto de la derivada y finalmente estableciendo la conexión entre integral y derivada.

2.3 Reglas fundamentales de integración

Las propiedades básicas de la integral se siguen directamente de su definición como el límite de una suma:

$$\int_a^b f(x)\, dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(\xi_i)\Delta x_i,$$

donde el intervalo $[a, b]$ es cortado en subintervalos o celdas de longitud Δx_i, el número ξ_i es cualquier valor en el i-ésimo subintervalo y se requiere que el máximo Δx_i tienda a cero conforme $n \to \infty$.

a. Aditividad

Sea c cualquier valor entre a y b. Si las integrales se interpretan como áreas y se recuerda que el área de una región que consiste de varias secciones es la suma de las áreas de las secciones (Fig. 2.14), esto nos conduce a la regla

(4)
$$\int_a^b f(x)\, dx = \int_a^c f(x)\, dx + \int_c^b f(x)\, dx.$$

Para una demostración analítica las subdivisiones se eligen de tal manera que el punto c aparezca como un punto de división, digamos $c = x_m$ (donde m varía con n). Entonces,

$$\sum_{i=1}^{n} f(\xi_i)\Delta x_i = \sum_{i=1}^{m} f(\xi_i)\Delta x_i + \sum_{i=m+1}^{n} f(\xi_i)\Delta x_i,$$

donde la primera suma en el miembro derecho corresponde a una subdivisión del intervalo $[a, c]$ en m celdas, y la segunda suma corresponde a una subdivisión del intervalo $[c, b]$. Ahora para $n \to \infty$ se obtiene nuestra regla para integrales.

Hasta ahora solamente hemos definido $\int_a^b f(x)\, dx$ cuando $a < b$. Para $a = b$, o bien para $a > b$, definimos la integral de tal manera que la regla de aditividad sea preservada. Por eso, para $c = a$ debe definirse

(5)
$$\int_a^a f(x)\, dx = 0,$$

y entonces para $b = a$ se sigue que

$$\int_a^c f(x)\,dx + \int_c^a f(x)\,dx = \int_a^a f(x)\,dx = 0.$$

Esto nos conduce a *definir* $\int_a^c f(x)\,dx$ para $c < a$ por la fórmula

$$\int_a^c f(x)\,dx = -\int_c^a f(x)\,dx, \tag{6}$$

donde el miembro derecho posee el significado establecido originalmente. Su *significado geométrico es que el área* bajo la curva $y = f(x)$ ha de ser

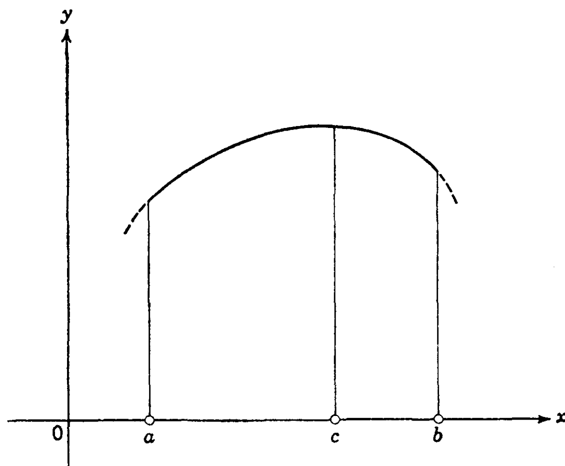


Figura 2.14.

considerada *negativa* si la dirección de movimiento desde el límite inferior de integración hacia el límite superior es el de las x decrecientes. Una ojeada a los ejemplos anteriores de integrales confirma que, en efecto, un intercambio en los límites de integración a y b da como resultado el cambio de signo en el valor de la integral.

b. Integral de una suma y de un producto con una constante

Si $f(x)$ y $g(x)$ son dos funciones cualesquiera (integrables), las leyes básicas de operar con límites implican

$$\begin{aligned}\int_a^b f(x)\,dx + \int_a^b g(x)\,dx &= \lim_{n\to\infty}\left[\sum_{i=1}^n f(\xi_i)\,\Delta x_i\right] + \lim_{n\to\infty}\left[\sum_{i=1}^n g(\xi_i)\,\Delta x_i\right] \\ &= \lim_{n\to\infty}\left[\sum_{i=1}^n f(\xi_i)\,\Delta x_i + \sum_{i=1}^n g(\xi_i)\,\Delta x_i\right] \\ &= \lim_{n\to\infty}\left\{\sum_{i=1}^n [f(\xi_i) + g(\xi_i)]\,\Delta x_i\right\};\end{aligned}$$

y, por lo tanto, la importante regla para la suma de dos funciones

$$\int_a^b f(x)\,dx + \int_a^b g(x)\,dx = \int_a^b [f(x) + g(x)]\,dx; \tag{7}$$

y análogamente para la diferencia:

$$\int_a^b f(x)\,dx - \int_a^b g(x)\,dx = \int_a^b [f(x) - g(x)]\,dx.$$

Además, con cualquier constante α es

$$\int_a^b \alpha f(x)\,dx = \lim_{n\to\infty} \sum_{i=1}^{n} \alpha f(\xi_i)\,\Delta x_i$$
$$= \alpha \lim_{n\to\infty} \sum_{i=1}^{n} f(\xi_i)\,\Delta x_i,$$

y así

$$\int_a^b \alpha f(x)\,dx = \alpha \int_a^b f(x)\,dx.$$

Las últimas dos reglas permiten integrar *"combinaciones lineales"* de dos o más funciones que pueden ser integradas individualmente. Así, para cualquier función cuadrática $y = Ax^2 + Bx + C$ con cualesquiera constantes A, B, C, se tiene

$$\int_a^b (Ax^2 + Bx + C)\,dx = \int_a^b Ax^2\,dx + \int_a^b Bx\,dx + \int_a^b C\,dx$$
$$= A\int_a^b x^2\,dx + B\int_a^b x\,dx + C\int_a^b 1\,dx$$
$$= \frac{A}{3}(b^3 - a^3) + \frac{B}{2}(b^2 - a^2) + C(b - a).$$

De la misma manera se integra el *polinomio* general

$$y = A_0x^n + A_1x^{n-1} + \cdots + A_{n-1}x + A_n:$$
$$\int_a^b y\,dx = \frac{1}{n+1}A_0(b^{n+1} - a^{n+1}) + \frac{1}{n}A_1(b^n - a^n) + \cdots$$
$$+ \tfrac{1}{2}A_{n-1}(b^2 - a^2) + A_n(b - a).$$

c. Estimación de integrales

Otra observación obvia concerniente a integrales es básica. Considérese para $a < b$ una función $f(x)$ la cual es positiva o cero en cada punto del intervalo $[a, b]$. Entonces,

$$\int_a^b f(x)\,dx \geq 0. \tag{9}$$

Esto surge inmediatamente si se escribe la integral como el límite de una suma y se observa que las sumas contienen solamente términos no negativos.

Más generalmente, si se tienen dos funciones f y g con la propiedad $f(x) \geq g(x)$ para todo x en el intervalo $[a, b]$ entonces

$$\int_a^b f(x)\,dx \geq \int_a^b g(x)\,dx. \tag{10}$$

Pues se tiene

$$\int_a^b f(x)\,dx - \int_a^b g(x)\,dx = \int_a^b [f(x) - g(x)]\,dx \geq 0,$$

puesto que $f(x) - g(x)$ jamás es negativa.

Aplíquese este resultado a una función $f(x)$ la cual es continua en el intervalo $[a, b]$. Sea M el máximo valor y m el mínimo valor de f en ese intervalo. Puesto que

$$m \leq f(x) \leq M$$

para todo x en $[a, b]$, se tiene

$$\int_a^b m\,dx \leq \int_a^b f(x)\,dx \leq \int_a^b M\,dx.$$

Recordando que para cualquier constante C

$$\int_a^b C\,dx = C\int_a^b 1\,dx = C(b - a),$$

se obtiene la desigualdad

$$m(b - a) \leq \int_a^b f(x)\,dx \leq M(b - a), \tag{11}$$

la cual da cotas superiores o inferiores sencillas para la integral definida de cualquier función continua.

Nuevamente, esta estimación es intuitivamente obvia. Si se considera la integral interpretada como un área, las cantidades $M(b - a)$ y $m(b - a)$ representan áreas de un rectángulo circunscripto y un rectángulo inscripto sobre la base común de longitud $b - a$ (véase Fig. 2.15).

d. El teorema del valor medio para integrales

La integral como un valor medio

Es significativa una interpretación ligeramente diferente de nuestras desigualdades en términos del ***promedio de la función f en un intervalo*** $[a, b]$. Para un número finito de cantidades $f_1, f_2, \ldots, f_n$, el ***promedio*** o ***media aritmética*** es el número

$$\frac{f_1 + f_2 + \cdots + f_n}{n}$$

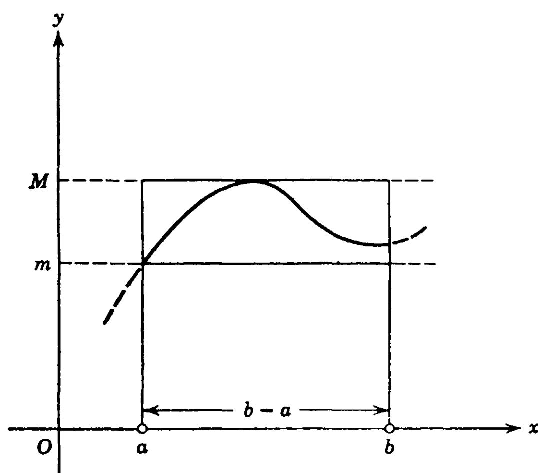


Figura 2.15.

Si se desea asignar un significado al valor promedio de un número infinito de cantidades $f(x)$ correspondientes a x arbitrarios en el intervalo $[a, b]$, es natural escoger primero un número finito n de valores de f, digamos $f(x_1), f(x_2), \ldots, f(x_n)$, para formar su promedio

$$\frac{f(x_1) + \cdots + f(x_n)}{n},$$

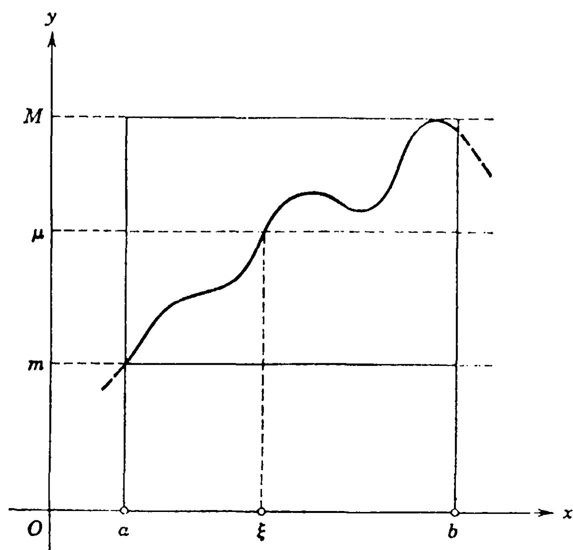


Figura 2.16. El valor medio μ de una función.

y entonces tomar el límite cuando n crece más allá de toda cota. El valor de este límite, si existe, dependerá mucho de cómo los puntos x_i se encuentren espaciados en el intervalo $[a, b]$. Un valor definido para el promedio de f se logra si se toman para los x_i los puntos que se obtienen cuando se divide el intervalo $[a, b]$ en n partes *iguales* de longitud $\Delta x_i = (b - a)/n$. Se tiene entonces

$$\frac{f(x_1) + \cdots + f(x_n)}{n} = \frac{1}{b - a} \sum_{i=1}^{n} f(x_i) \Delta x_i,$$

y es claro que en el límite, para $n \to \infty$, los n-ésimos promedios convergen hacia el valor

$$\mu = \frac{1}{b - a} \int_a^b f(x)\, dx = \frac{\int_a^b f(x)\, dx}{\int_a^b dx}.$$

A μ se le llamará la "media aritmética" o bien el *valor medio* de f en el intervalo $[a, b]$. Nuestras desigualdades entonces establecen simplemente que el valor medio de una función continua no puede ser mayor que el valor máximo ni menor que el valor mínimo de la función (Fig. 2.16).

Puesto que la función $f(x)$ es continua en el intervalo $[a, b]$, deben existir puntos en el intervalo donde f posee el valor M o el valor m. Por el teorema del valor intermedio para funciones continuas, deberá entonces existir también un punto ξ en el intervalo donde f tome, en efecto, el valor intermedio μ. Se ha probado entonces:

TEOREMA DEL VALOR MEDIO. *Para una función continua $f(x)$ en el intervalo $[a, b]$, existe un valor ξ en el intervalo tal que*

$$\int_a^b f(x)\, dx = f(\xi)(b - a). \tag{12}$$

Este es el sencillo pero muy importante *teorema del valor medio del cálculo integral.* En palabras, afirma que el valor medio de una función continua en un intervalo pertenece al rango de la función.

El teorema asegura solamente la *existencia* de por lo menos un ξ en el intervalo, para el cual $f(\xi)$ es igual al valor promedio de f, pero no da ninguna información adicional sobre la ubicación de ξ.

Nótese que la fórmula que expresa al teorema del valor medio permanece válida si los límites a y b son intercambiados; por lo tanto, el teorema del valor medio es correcto también cuando $a > b$.

El teorema del valor medio generalizado. En vez de la sencilla media aritmética es necesario considerar a menudo "medias pesadas" de n cantidades $f_1, \ldots, f_n$, dadas por

$$\frac{p_1 f_1 + p_2 f_2 + \cdots + p_n f_n}{p_1 + p_2 + \cdots + p_n} = \mu,$$

donde "los factores peso" p_i son cualesquiera cantidades positivas. Si, por ejemplo, $p_1, p_2, \ldots$, son, realmente, los pesos de partículas localizadas en los puntos respectivos $f_1, f_2, \ldots, f_n$ del eje x, entonces μ representará la posición del centro de gravedad. Si todos los pesos p_i son iguales, la cantidad μ es justamente la media *aritmética* definida anteriormente.

Para una función $f(x)$ puede formarse análogamente la media pesada

$$\mu = \frac{\int_a^b f(x) p(x)\, dx}{\int_a^b p(x)\, dx} \tag{13}$$

sobre el intervalo $[a, b]$ donde $p(x)$, la *función peso,* es cualquier función positiva en el intervalo. La hipótesis de que p es positiva garantiza que el denominador no se anula.

La media pesada μ está situada también entre el máximo valor M y el mínimo valor m de la función f en el intervalo.

En efecto, multiplicando la desigualdad

$$m \leq f(x) \leq M,$$

por el número *positivo* $p(x)$, se encuentra que

$$mp(x) \leq f(x)p(x) \leq Mp(x).$$

La integración conduce entonces a

$$m \int_a^b p(x)\, dx \leq \int_a^b f(x)p(x)\, dx \leq M \int_a^b p(x)\, dx.$$

Dividiendo por la cantidad positiva $\int_a^b p(x)\, dx$, indudablemente se obitene el resultado

$$m \leq \mu \leq M.$$

Si aquí $f(x)$ es continua, se concluye del teorema del valor intermedio (p. 68) que $\mu = f(\xi)$, donde ξ es un valor apropiado en el intervalo $a \leq \xi \leq b$. Esto conduce al siguiente *teorema del valor medio generalizado del cálculo integral:*

Si $f(x)$ y $p(x)$ son continuas en el intervalo $[a, b]$ y además $p(x)$ es positiva en este intervalo, entonces existe un valor ξ en el intervalo tal que

$$\int_a^b f(x)p(x)\, dx = f(\xi) \int_a^b p(x)\, dx. \tag{14}$$

El caso especial $p(x) = 1$ conduce al teorema del valor medio anterior.

2.4 La integral como función del límite superior (integral indefinida)

Definición y fórmula básica

El valor de la integral de una función $f(x)$ depende de los límites de integración a y b: La integral es una función de los dos límites a y b. Para estudiar más de cerca esta dependencia sobre los límites imaginemos que el límite inferior es un número fijo, digamos α, denotemos la variable de integración ya no por x sino por u (véase p. 148) y denótese el límite superior por x en vez de denotarlo por b, de manera de indicar que se considerará al límite superior como la variable y que se desea investigar el valor de la integral como función de su límite superior. En efecto, escribimos

$$\phi(x) = \int_{\alpha}^{x} f(u)\, du.$$

La función $\phi(x)$ se denomina una *integral indefinida* de la función $f(x)$. Cuando se habla de *una* y no de *la* integral indefinida, se sugiere que en vez del límite inferior α pudo haberse elegido cualquier otro, en cuyo caso se obtendría ordinariamente un valor diferente para la integral. Geométrica-

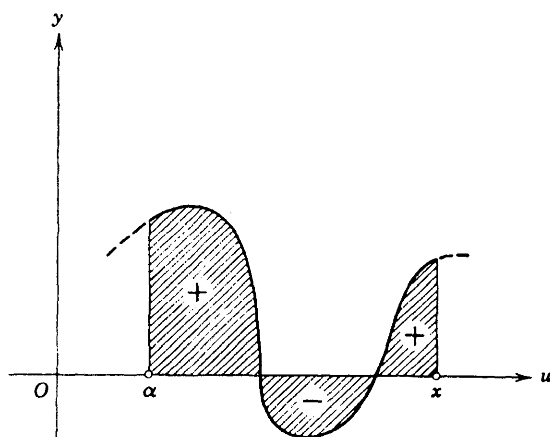


Figura 2.17 La integral indefinida como un área.

mente, la integral indefinida $\phi(x)$ está dada por el área (mostrada por el sombreado en la Fig. 2.17) bajo la curva $y = f(u)$ y acotada por el eje u, la ordenada $u = \alpha$ y la variable ordenada $u = x$, estando el signo determinado por las reglas discutidas anteriormente (p. 148).

Cualquier integral *definida* particular se encuentra a partir de la integral indefinida $\phi(x)$. Sin duda, por las reglas básicas para integrales,

$$\int_a^b f(u)\,du = \int_a^\alpha f(u)\,du + \int_\alpha^b f(u)\,du$$
$$= -\int_\alpha^a f(u)\,du + \int_\alpha^b f(u)\,du = \phi(b) - \phi(a).$$

En particular, puede expresarse cualquier otra integral indefinida con límite inferior α' en términos de $\phi(x)$:

$$\int_{\alpha'}^x f(u)\,du = \phi(x) - \phi(\alpha').$$

Como se ve, *cualquier integral indefinida difiere de la integral indefinida especial* $\phi(x)$ *sólo por una constante.*

Continuidad de la integral indefinida

Si la función $f(x)$ es continua en el intervalo $[a, b]$ y α es un punto de este intervalo, entonces la integral indefinida

$$\phi(x) = \int_\alpha^x f(u)\,du$$

representa una función de x la cual está definida nuevamente en el mismo intervalo. Como se ve fácilmente: *La integral indefinida* $\phi(x)$ *de una función continua* $f(x)$ *es asimismo continua.* En efecto, si x e y son dos valores cualesquiera en el intervalo, se tiene por el teorema del valor medio que

$$(15) \qquad \phi(y) - \phi(x) = \int_x^y f(u)\,du = f(\xi)(y - x)$$

donde ξ es algún valor en el intervalo con puntos extremos x e y. De la continuidad de f se tiene entonces

$$\lim_{y\to x} \phi(y) = \lim_{y\to x} [\phi(x) + f(\xi)(y - x)] = \phi(x) + f(x)\cdot 0 = \phi(x),$$

lo que muestra que ϕ es continua. Más específicamente, en cualquier intervalo cerrado se tiene $|\phi(y) - \phi(x)| \leq M\,|y - x|$, donde M es el máximo de $|f|$ en el intervalo, de modo que ϕ es aún Lipschitz-continua.

La fórmula (15) para $\phi(y) - \phi(x)$ muestra: que $\phi(x)$ es una función creciente de x en el caso en que f sea positiva en todo el intervalo, o sea, para $y > x$

$$\phi(y) = \phi(x) + f(\xi)(y - x) > \phi(x).$$

El formar la integral indefinida de una función es un camino importante para *general nuevos tipos de funciones.* En la sección 2.5 se aplicará este método para introducir la función logaritmo. Esto nos hará también vislumbrar el hecho de que teoremas generales del análisis matemático conducen a las más notables fórmulas específicas.

Como se verá en la sección 3.14a (p. 298), la definición de nuevas funciones por medio de integrales de funciones ya definidas es un procedimiento satisfactorio si se desean establecer las definiciones (por ejemplo, de las funciones trigonométricas) sobre una base puramente analítica en vez de atenerse a explicaciones geométricas intuitivas.

2.5 El logaritmo definido mediante una integral

a. Definición de la función logaritmo

En la sección 2.2 se tuvo buen éxito en expresar $\int_a^b x^\alpha\, dx$ para cualquier racional $\alpha \neq -1$ en términos de potencias de a y de b. Para $\alpha = -1$ se pudo solamente representar la integral como el límite de la sucesión

$$\int_a^b \frac{1}{u}\, du = \lim_{n\to\infty} n(\sqrt[n]{b/a} - 1).$$

Independientemente de las discusiones de la sección 2.2 se *introduce* ahora la función representada por la integral indefinida

$$\int_1^x \frac{1}{u}\, du,$$ [1]

o bien, geométricamente, mediante el área bajo una hipérbola como se indica en la Fig. 2.18. La denominamos el *logaritmo* de x, o bien, más precisamente, el *logaritmo natural* de x, y se escribe

$$\log x = \int_1^x \frac{1}{u}\, du. \tag{16}$$

Puesto que $y = 1/u$ es una función continua y positiva para todo $u > 0$, la función $\log x$ está definida para todo $x > 0$, siendo además continua y también monótona creciente. La elección de 1 como el límite inferior en la integral indefinida para $\log x$ es cuestión de conveniencia. Implica que

$$\log 1 = 0, \tag{17}$$

[1] En esta sección otra vez se utilizará libremente el hecho de que la integral de una función continua (en este caso la función $1/u$) existe; la demostración general está dada en el Suplemento.

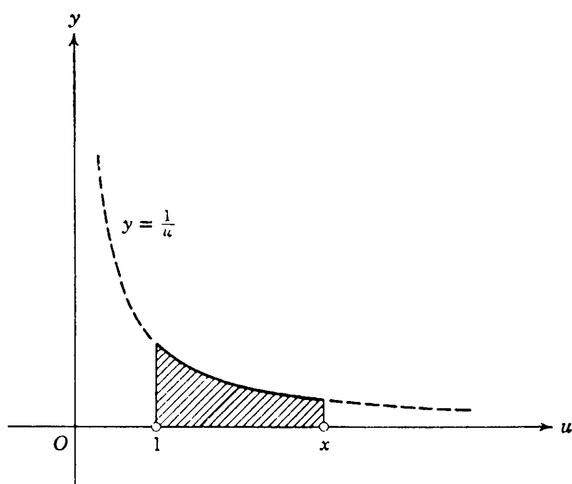


Figura 2.18 Log x representado por un área.

y que $\log x$ es positivo para $x > 1$ y negativo para x entre cero y 1 (Fig. 2.19). Cualquier integral definida de $1/u$ entre límites positivos a y b puede ser expresada en términos de logaritmos mediante la fórmula (véase pp. 165-166)

$$\int_a^b \frac{1}{u}\,du = \log b - \log a. \tag{18}$$

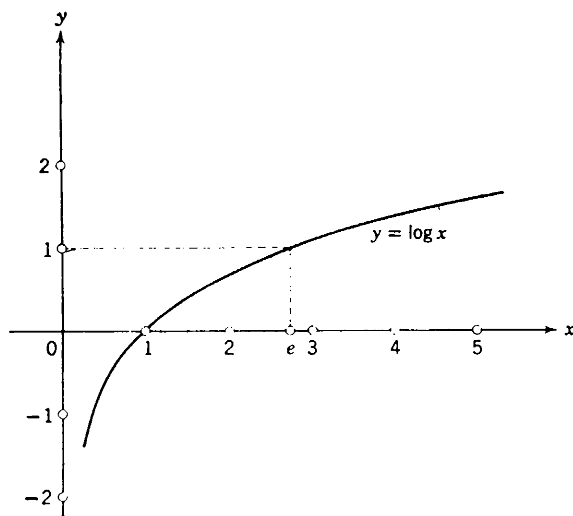


Figura 2.19 El logaritmo natural.

Geométricamente, esta integral representa el área bajo la hipérbola $y = 1/x$ entre las ordenadas $x = a$ y $x = b$.

b. El teorema de adición para logaritmos

La propiedad fundamental que justifica el nombre tradicional para $\log x$ está expresada por el

TEOREMA DE ADICIÓN. *Para todo x e y positivos*

$$\log(xy) = \log x + \log y. \tag{19}$$

DEMOSTRACIÓN. El teorema de adición se escribe en la forma

$$\log(xy) - \log y = \log x,$$

o bien

$$\int_y^{xy} \frac{1}{v}\,dv = \int_1^x \frac{1}{u}\,du,$$

donde deliberadamente se han elegido diferentes letras para las variables de integración en las dos integrales. La igualdad de las dos integrales se deducirá del hecho de que las sumas de aproximación poseen el mismo valor para elecciones adecuadas de subdivisiones y de puntos intermedios. Supóngase primero que $x > 1$. Entonces

$$\int_1^x \frac{1}{u}\,du = \lim_{n\to\infty} \sum_{i=1}^n \frac{1}{\xi_i}\,\Delta u_i,$$

donde $u_0 = 1$, u_1, u_2, . . . , $u_n = x$ representan los puntos que aparecen en una subdivisión del intervalo $[1, x]$ y ξ_i está situado en la i-ésima celda. Haciendo $v_i = yu_i$, $\eta_i = y\xi_i$, se ve que los puntos v_0, v_1, . . . , v_n corresponden a una subdivisión del intervalo $[y, xy]$ con puntos intermedios $\eta_i = \xi_i y$. Obviamente,

$$\Delta v_i = y\,\Delta u_i,$$

de modo que

$$\sum_{i=1}^n \frac{1}{\eta_i}\,\Delta v_i = \sum_{i=1}^n \frac{1}{\xi_i}\,\Delta u_i.$$

Para n tendiendo a infinito se obtiene la identidad deseada entre integrales para el caso $x > 1$.

Para $x = 1$ el teorema de adición se cumple trivialmente, puesto que $\log 1 = 0$. Para probar el teorema también para el caso $0 < x < 1$, se observa que entonces $1/x > 1$, y por tanto

$$\begin{aligned}\log x + \log y &= \log x + \log\left(\frac{1}{x}xy\right)\\ &= \log x + \log\frac{1}{x} + \log(xy)\\ &= \log\frac{1}{x} + \log x + \log(xy)\\ &= \log\left(\frac{1}{x}x\right) + \log(xy)\\ &= \log 1 + \log(xy) = \log(xy).\end{aligned}$$

Esto completa la demostración del teorema de adición.

Una demostración del teorema de adición puede también basarse en la fórmula (3) (p. 156), de acuerdo con la cual

$$\log x = \lim_{n\to\infty} n(\sqrt[n]{x} - 1).$$

Entonces

$$\begin{aligned}\log(xy) &= \lim_{n\to\infty} n(\sqrt[n]{xy} - 1)\\ &= \lim_{n\to\infty}[n(\sqrt[n]{x} - 1)\sqrt[n]{y} + n(\sqrt[n]{y} - 1)]\\ &= [\lim_{n\to\infty} n(\sqrt[n]{x} - 1)](\lim_{n\to\infty}\sqrt[n]{y}) + \lim_{n\to\infty} n(\sqrt[n]{y} - 1)\\ &= \log x + \log y,\end{aligned}$$

puesto que $\lim_{n\to\infty}\sqrt[n]{y} = 1$ (véase p. 87).

Aplicando el teorema de adición para el caso especial $y = 1/x$, éste conduce a

$$\log 1 = \log x + \log\frac{1}{x},$$

o bien

(20) $$\log\frac{1}{x} = -\log x.$$

Entonces, más generalmente,

(21) $$\log\frac{y}{x} = \log y + \log\frac{1}{x} = \log y - \log x.$$

Aplicando repetidamente el teorema de adición a un producto de n factores se tiene

$$\log(x_1 x_2 \cdots x_n) = \log x_1 + \log x_2 + \cdots + \log x_n.$$

En particular, se encuentra que para cualquier entero positivo n

(22) $$\log(x^n) = n\log x.$$

Esta identidad se cumple también para $n = 0$, puesto que $x^0 = 1$, y puede ser extendida a enteros negativos n observando que

$$\log(x^n) = \log\left(\frac{1}{x^{-n}}\right) = -\log(x^{-n}) = -(-n)\log x = n \log x.$$

Para cualquier racional $\alpha = m/n$ y cualquier a positivo puede formarse $a^\alpha = a^{m/n} = x$. Se tiene entonces

$$\log x = \frac{1}{n}\log x^n = \frac{1}{n}\log a^m = \frac{m}{n}\log a = \alpha \log a.$$

Así, la identidad

$$\log(a^\alpha) = \alpha \log a \tag{23}$$

se cumple para cualquier real positivo a y cualquier racional α.

2.6 Función exponencial y potencias

a. El logaritmo del número e

La constante e obtenida en la p. 102 como el límite de $(1 + 1/n)^n$ desempeña un marcado papel para la función $\log x$. Sin duda, el número e está caracterizado por la ecuación [1]

$$\log e = 1.$$

Para la demostración, se observa que la continuidad de la función $\log x$ implica

$$\log e = \log\left[\lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n\right] = \lim_{n\to\infty}\log\left[\left(1 + \frac{1}{n}\right)^n\right]$$
$$= \lim_{n\to\infty} n \log\left(1 + \frac{1}{n}\right).$$

Ahora, por el teorema del valor medio del cálculo integral,

$$\log\left(1 + \frac{1}{n}\right) = \int_1^{1+1/n}\frac{1}{u}\,du = \frac{1}{\xi}\frac{1}{n},$$

donde ξ es algún número entre 1 y $1 + 1/n$ que depende de la elección de n. Obviamente, $\lim_{n\to\infty}\xi = 1$, de modo que

$$\log e = \lim_{n\to\infty}\frac{1}{\xi} = 1. \tag{24}$$

[1] Esto significa geométricamente que el área acotada por la hipérbola $y = 1/x$ y las rectas $y = 0$, $x = 1$ y $x = e$, tiene el valor uno (véase Fig. 2.18).

b. La función inversa del logaritmo. La función exponencial

De la relación $\log e = 1$ se sigue que para todo racional α es

$$\log (e^\alpha) = \alpha \log e = \alpha.$$

Esto muestra que todo número racional α aparece como un valor del $\log x$ para algún x positivo. Puesto que $\log x$ es continuo, toma entonces cualquier valor entre dos números racionales, esto es, todos los valores reales contenidos entre tales números. Se sigue que cuando x varía sobre todos los valores positivos, los valores de $y = \log x$ recorren todos los números y.

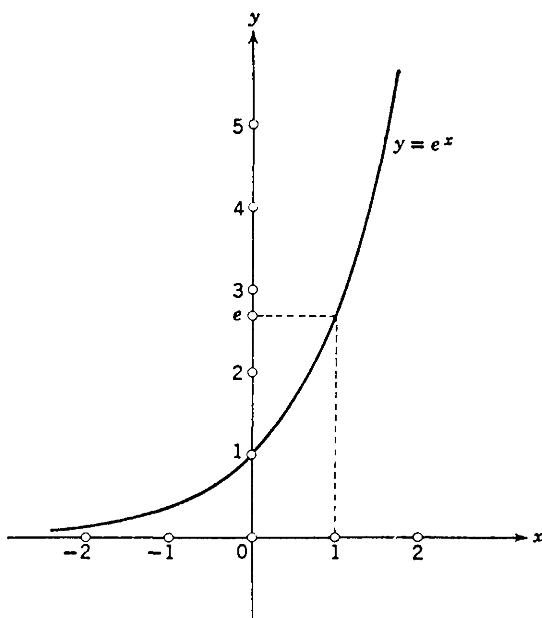


Figura 2.20 La función exponencial.

Puesto que $\log x$ es monótono creciente, existe para todo número real y exactamente un real x positivo tal que $\log x = y$. La solución x de la ecuación $y = \log x$ está dada por la función inversa del logaritmo, la cual se denotará por $x = E(y)$. Se sabe entonces que $E(y)$ (Fig. 2.20) está definida y es positiva para todo y, y que es también continua y creciente (véase pp. 65-66).

Puesto que las ecuaciones $y = \log x$ y $x = E(y)$ se cumplen para la misma relación entre x e y, la ecuación $\alpha = \log (e^\alpha)$, la cual es válida para α racional, puede escribirse también en la forma

$$E(\alpha) = e^\alpha.$$

Se ve que: para todo racional α el valor de $E(\alpha)$ es la α-ésima potencia del número e. Para $\alpha = m/n$ racional la potencia e^{α} está definida directamente como $\sqrt[n]{e^m}$. Para α irracional la expresión e^{α} está definida más naturalmente representando a α como el límite de una sucesión de números racionales α_n y haciendo $e^{\alpha} = \lim (e^{\alpha_n})$. Puesto que $e^{\alpha_n} = E(\alpha_n)$, y puesto que la función $E(y)$ depende continuamente de y, se puede estar seguro de que el límite de e^{α_n} existe y que posee el valor $E(\alpha)$, independientemente de la sucesión especial utilizada para aproximar a α. Esto prueba que la ecuación $E(\alpha) = e^{\alpha}$ se cumple también para α irracional. Para todo α real puede escribirse ahora e^{α} en vez de $E(\alpha)$. Denominamos a e^x *función exponencial.* Esta función está definida y es continua para todo x, y es creciente y positiva en todas partes.

Puesto que las ecuaciones $y = \log x$ y $x = e^y$ son dos formas de expresar la misma relación entre los números x e y, se ve que la designación $\log x$, el "logaritmo natural" de x (como se define aquí mediante una integral), se mantiene para el *logaritmo de base e,* con la interpretación que a ese término se le daría en matemáticas elementales; esto es, $\log x$ es el exponente de aquella potencia de e que es igual a x, o sea,

$$e^{\log x} = x. \tag{25}$$

Puede escribirse [1] $\log x = \log_e x$.

Análogamente, $x = e^y$ es aquel número cuyo logaritmo es y, o sea,

$$\log e^y = y. \tag{26}$$

Desde el punto de vista del cálculo, es realmente más fácil introducir en primer término logaritmos naturales como integrales de la función sencilla $y = 1/x$, como se hizo aquí, y definir después potencias de e tomando la inversa de la función logaritmo. De esta forma la continuidad y monotonía de las funciones $\log x$ y e^x aparecen justamente como consecuencia de teoremas generales y no requieren argumentos especiales.

c. La función exponencial como límite de potencias

Originalmente se obtuvo el número e como el límite

$$e = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n.$$

Una fórmula más general representa e^x para cualquier x como un límite, a saber:

---

[1] El lector puede pensar que el nombre "logaritmo natural" debió haber sido reservado más bien para logaritmos de base 10. Sin embargo, históricamente, la primera tabla de logaritmos, publicada por Napier en 1614, dio esencialmente logaritmos de base e. Los logaritmos de base 10 fueron introducidos sólo subsecuentemente por Briggs debido a sus obvias ventajas de cálculo.

(27) $$e^x = \lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n.$$

Para la demostración, es suficiente mostrar que la sucesión

$$s_n = \log\left(1 + \frac{x}{n}\right)^n$$

posee el límite x. Pues entonces la sucesión de valores

$$e^{s_n} = \left(1 + \frac{x}{n}\right)^n$$

debe tender a e^x puesto que la función exponencial es continua. Ahora bien,

$$s_n = n \log\left(1 + \frac{x}{n}\right) = n\int_1^{1+x/n} \frac{1}{\xi}\,d\xi.$$

Por el teorema del valor medio del cálculo integral se tiene

$$s_n = n\frac{1}{\xi_n}\left[\left(1 + \frac{x}{n}\right) - 1\right] = \frac{x}{\xi_n},$$

donde ξ_n es algún valor entre uno y $1 + x/n$. Puesto que obviamente ξ_n tiende a uno para n tendiendo a ∞, se tiene en efecto $\lim\limits_{n\to\infty} s_n = x$.

d. Definición de potencias arbitrarias de números positivos

Las potencias arbitrarias de cualesquiera números positivos pueden ahora ser expresadas en términos de las funciones exponencial y logaritmo.[1]

Se encuentra para α racional y cualquier x positivo que la relación

$$\log(x^\alpha) = \alpha \log x$$

se cumple. Se escribe esta ecuación en la forma

$$x^\alpha = e^{\alpha \log x}.$$

Para α irracional se representa a α nuevamente como el límite de una sucesión de números racionales α_n, y se define

$$x^\alpha = \lim_{n\to\infty} x^{\alpha_n} = \lim_{n\to\infty} e^{\alpha_n \log x}.$$

La continuidad de la función exponencial implica nuevamente que el límite existe y que posee el valor $e^{\alpha \log x}$, puesto que

$$e^{\alpha \log x} = e^{\lim(\alpha_n \log x)} = \lim e^{\alpha_n \log x}.$$

[1] Esto evita la definición "elemental", más incómoda, y la justificación de estos procesos mediante el paso al límite a partir de exponentes racionales, indicado en la p. 109.

Por lo tanto, la ecuación

$$x^{\alpha} = e^{\alpha \log x} \tag{28}$$

se cumple de manera completamente general para todo α y cualquier x positivo. Haciendo $\log x = \beta$, o bien, lo que es lo mismo, $x = e^{\beta}$, se infiere que

$$(e^{\beta})^{\alpha} = e^{\alpha\beta}, \tag{29}$$

y, más generalmente, para todo x positivo:

$$(x^{\alpha})^{\beta} = (e^{\alpha \log x})^{\beta} = e^{\alpha\beta \log x} = x^{\alpha\beta}.$$

Otra regla para operar con potencias, que es fácilmente establecida en forma completamente general, es la *ley de multiplicación,* o sea,

$$x^{\alpha}x^{\beta} = x^{\alpha+\beta},$$

donde x es un número positivo y α y β son arbitrarios. Es suficiente probar la correspondiente fórmula obtenida tomando los logaritmos de ambos lados:

$$\log (x^{\alpha}x^{\beta}) = \log (x^{\alpha+\beta}).$$

Ahora bien, por las reglas (19), (26) y (28) ya establecidas, se sigue que

$$\begin{aligned}\log (x^{\alpha}x^{\beta}) &= \log x^{\alpha} + \log x^{\beta} = \log (e^{\alpha \log x}) + \log (e^{\beta \log x})\\ &= \alpha \log x + \beta \log x = (\alpha + \beta) \log x\\ &= \log (e^{(\alpha+\beta) \log x}) = \log (x^{\alpha+\beta}).\end{aligned}$$

e. Logaritmos de base arbitraria

Es fácil expresar logaritmos relativos a una base diferente de e en términos de logaritmos naturales. Si, para un número positivo, a, la ecuación $x = a^y$ se satisface, se escribe entonces

$$y = \log_a x.$$

Ahora bien, $a^y = e^{y \log a}$, de modo que $x = e^{y \log a}$, o sea que $y \log a = \log x$. Se sigue que

$$\log_a x = \frac{\log x}{\log a}, \tag{30}$$

donde $\log x$ es el logaritmo natural de base e. En particular, los logaritmos comunes de base 10 están dados por

$$\log_{10} x = \frac{\log x}{\log 10}.$$

Puesto que los logaritmos de cualquier base a son proporcionales a los logaritmos naturales, satisfacen el mismo teorema de adición:

$$\log_a x + \log_a y = \log_a (xy).$$

2.7 La integral de una potencia arbitraria de x

En la sección 2.2 se obtuvo la fórmula

$$\int_a^b u^\alpha \, du = \frac{b^{\alpha+1} - a^{\alpha+1}}{\alpha + 1},$$

para todo racional $\alpha \neq -1$. (El caso $\alpha = -1$ se vio que conduce al logaritmo.) Para evaluar la integral cuando α es un número irracional es suficiente discutir la integral indefinida

$$\phi(x) = \int_1^x u^\alpha \, du,$$

de la cual todas las integrales definidas con límites positivos a y b pueden ser obtenidas. Supóngase $x > 1$ (el caso $x < 1$ puede ser tratado de la misma manera después de intercambiar los límites). Se tiene entonces, por (28), que

$$u^\alpha = e^{\alpha \log u},$$

donde $\log u \geq 0$ para u en el intervalo de integración. Sean β y γ dos números racionales cualesquiera diferentes de -1 para los cuales

$$\beta \leq \alpha \leq \gamma.$$

Entonces también

$$\beta \log u \leq \alpha \log u \leq \gamma \log u.$$

Puesto que la función exponencial es creciente, esto implica

$$e^{\beta \log u} \leq e^{\alpha \log u} \leq e^{\gamma \log u};$$

esto es,

$$u^\beta \leq u^\alpha \leq u^\gamma.$$

Se tiene entonces

$$\int_1^x u^\beta \, du \leq \phi(u) \leq \int_1^x u^\gamma \, du.$$

Las integrales de u^β y u^γ se evaluaron anteriormente, conduciendo a

$$\frac{1}{\beta + 1}(x^{\beta+1} - 1) \leq \phi(x) \leq \frac{1}{\gamma + 1}(x^{\gamma+1} - 1).$$

Si se hacen ahora converger los números racionales β y γ a α, se obtiene en el límite

$$\phi(x) = \frac{1}{\alpha + 1}(x^{\alpha+1} - 1),$$

puesto que $x^{\beta+1} = e^{(\beta+1)\log x}$ y $x^{\gamma+1} = e^{(\gamma+1)\log x}$ tienden a $e^{(\alpha+1)\log x} = x^{\alpha+1}$ debido a la continuidad de la función exponencial. El mismo resultado se sigue para x entre cero y uno. Así, de manera general, para a, b positivos es

$$\int_a^b u^\alpha \, du = \phi(b) - \phi(a) = \frac{1}{\alpha + 1}(b^{\alpha+1} - a^{\alpha+1}),$$

tal como para α racional.

Cuando α es un entero positivo la fórmula permanece válida aun cuando los límites a o b se tornen cero o bien negativos, y es fácil extender la fórmula directamente a esos casos.

2.8 La derivada

El concepto de derivada, análogamente al de integral, posee un inmediato origen intuitivo y es fácil de entender. Además, abre la puerta hacia un caudal enorme de hechos matemáticos e ideas. Y el estudiante sólo gradualmente se hará conocedor de la variedad de aplicaciones significativas y de la potencia de las técnicas que se desarrollarán en este libro.

El concepto de derivada es sugerido primero mediante la noción intuitiva de la *tangente a una curva suave* $y = f(x)$ en un punto P de coordenadas x e y. Esta tangente es caracterizada por el ángulo α entre su dirección y el eje x positivo. Pero, ¿cómo obtiene uno este ángulo a partir de la descripción analítica de la función $f(x)$? El conocimiento de los valores de x e y en el punto P no basta para determinar el ángulo α, puesto que hay un número infinito de rectas, aparte de la tangente que pasa por P. Por otra parte, para determinar α no se requiere conocer la función $f(x)$ en su comportamiento global; el conocimiento de la función en una vecindad arbitraria del punto P debe ser suficiente para determinar la dirección α, no importa cuán minúscula sea elegida esa vecindad. Esto indica que se debería definir la dirección de la tangente a una curva $y = f(x)$ mediante un proceso de límite, como se hará en breve.

El problema de calcular la dirección de tangentes, o de "derivación", fue sugerido a los matemáticos alrededor del siglo XVI por problemas de optimización, esto es, cuestiones de máximos y mínimos que aparecen en la geometría, mecánica y óptica. (Véase la discusión en la sección 3.6.)

Otro problema de capital importancia que conduce a la derivación es el de dar un significado matemático preciso a la noción intuitiva de *velocidad* en un movimiento arbitrario no uniforme (véase p. 183).

Se principiará con el problema de describir analíticamente la tangente a una curva mediante un proceso de límite.

a. La derivada y la tangente

Definición geométrica. De acuerdo con la intuición ingenua, se define la tangente a la curva dada, $y = f(x)$, en uno de sus puntos, P, por medio de los siguientes procesos de límite geométricos (Fig. 2.21). Considérese un segundo punto P_1 sobre la curva cercano a P. Por los dos puntos P y P_1 se traza una línea recta secante de la curva. Si ahora el punto P_1 se mueve a lo largo de la curva hacia el punto P, entonces es de esperarse que la secante se aproxime a una posición límite la cual es independiente del lado del cual P_1 tiende a P. Esta *posición límite de la secante es la tangente;* la aseveración de que existe tal posición límite de la secante es equivalente a la suposición de que la curva posee una tangente definida o bien una dirección definida en el punto P. Se ha utilizado la palabra "suposición" debido a que en realidad se ha hecho una. La hipótesis de que la tangente existe en cada punto no es de ninguna manera cierta para todas las curvas que representan funciones sencillas. Por ejemplo, cualquier curva con una esquina o vértice en un punto P no posee ahí una dirección determinada de manera única; por ejemplo, la curva definida por $y = |x|$, en $(0, 0)$. (Véase la discusión en las pp. 187-188.)

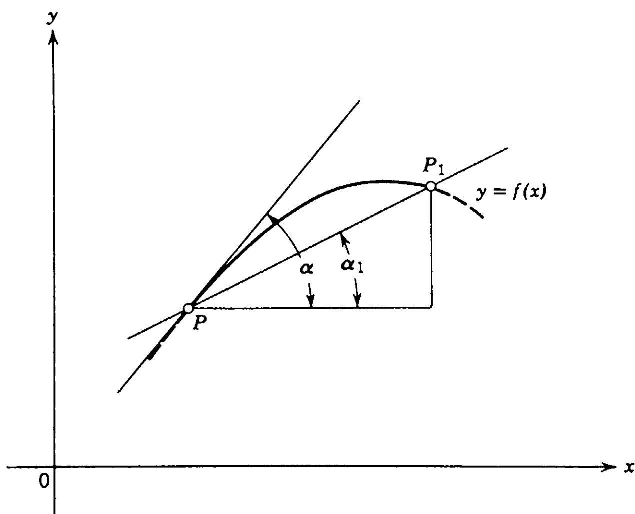


Figura 2.21 Secante y tangente.

Puesto que nuestra curva está representada por medio de una función $y = f(x)$, debe formularse el proceso de límite geométrico analíticamente con referencia a $f(x)$. Este proceso de límite analítico se denomina *derivación* de $f(x)$.

Considérese al ángulo que una línea recta forma con el eje x como aquél según el cual el eje x positivo debe ser girado en la dirección positiva o en el sentido contrario al de las manecillas del reloj[1] de modo que quede por primera vez paralelo a la recta. (Este sería un ángulo α en el intervalo $0 \leq \alpha \leq \pi$.) Sea α_1 el ángulo que la secante PP_1 forma con el eje x positivo (ver Fig. 2.22) y α el ángulo que la tangente forma con el eje x positivo. Entonces

$$\lim_{P_1 \to P} \alpha_1 = \alpha,$$

donde el significado de los símbolos es obvio. Sean x, y y x_1, y_1 las coordenadas de los puntos P y P_1 respectivamente. Entonces se tiene de inmediato.[2]

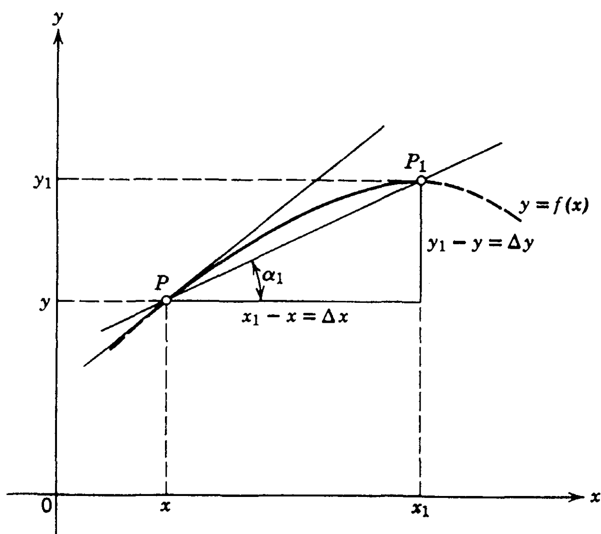


Figura 2.22.

$$\tan \alpha_1 = \frac{y_1 - y}{x_1 - x} = \frac{f(x_1) - f(x)}{x_1 - x};$$

así, nuestro proceso de límite (omitiendo el caso $\alpha = \pi/2$ de una tangente perpendicular) está representado por la ecuación

$$\lim_{x_1 \to x} \frac{f(x_1) - f(x)}{x_1 - x} = \lim_{x_1 \to x} \tan \alpha_1 = \tan \alpha.$$

---

1 Esto es, en una dirección tal que una rotación de $\pi/2$ lo hace coincidir con el eje positivo y.

2 Para que esta ecuación pueda tener significado debe suponerse que tanto x como x_1 están situadas en el dominio de f. En lo que sigue, tácitamente se harán las hipótesis correspondientes en los pasos que dan lugar a procesos de límite.

Notación. La expresión

$$\frac{f(x_1) - f(x)}{x_1 - x} = \frac{y_1 - y}{x_1 - x} = \frac{\Delta y}{\Delta x}$$

se denomina *cociente incremental* de la función $y = f(x)$, y los símbolos Δy y Δx denotan las diferencias de la función $y = f(x)$ y de la variable independiente x. (Aquí, como en la p. 146, el símbolo Δ es una abreviación para diferencia y *no* es un factor.) La tangente trigonométrica de α, la "pendiente" de la curva,[1] es por eso igual al límite al cual el cociente incremental de nuestra función tiende cuando x_1 tiende a x.

A este límite del cociente incremental se le denomina *derivada*[2] de la función $y = f(x)$ en el punto x. Generalmente se utilizará la notación de Lagrange, $y' = f'(x)$, para denotar la derivada, o bien, como lo hiciera Leibnitz, el símbolo[3] dy/dx, o $df(x)/dx$, o $(d/dx)f(x)$. En la p. 192 se discutirá el significado de la notación de Leibnitz con más detalle. Aquí se puntualiza que: La notación $f'(x)$ indica el hecho de que la *derivada es, ella misma, una función de x,* puesto que un valor de $f'(x)$ corresponde a cada valor de x en el intervalo considerado. Este hecho es destacado en ocasiones mediante el uso de los términos *función derivada, curva derivada.* La definición de la derivada aparece en varias formas diferentes:

$$f'(x) = \lim_{x_1 \to x} \frac{f(x_1) - f(x)}{x_1 - x} = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h},$$

donde en la segunda expresión x_1 se reemplaza por $x + h$; o bien en la notación de Leibnitz,

$$\frac{dy}{dx} = \frac{df(x)}{dx} = f'(x) = \lim_{x_1 \to x} \frac{f(x_1) - f(x)}{x_1 - x} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}.$$

Si f está definida en una vecindad del punto x, entonces el cociente $[f(x + h) - f(x)]/h$ está definido como una función de h para todos los valores $h \neq 0$ para los cuales $|h|$ es lo suficientemente pequeño para asegurar que $x + h$ está en el intervalo bajo consideración. La definición de $f'(x)$ como límite requiere que $\left| \frac{f(x + h) - f(x)}{h} - f'(x) \right|$ sea arbitrariamente pequeño para todo $h \neq 0$ (positivo *o* negativo) tal que $|h|$ sea suficientemente pequeño.

Cálculo analítico de derivadas. El concepto intuitivo y la noción analítica general de derivada son simples y naturales. Menos obvio es el procedimiento de llevar a cabo verdaderamente tales procesos de límite.

---

[1] Las palabras *gradiente* o bien *coeficiente de dirección* son utilizadas ocasionalmente.

[2] El término *coeficiente diferencial* es también utilizado en textos antiguos.

[3] La notación de Cauchy $Df(x)$ y la notación de Newton $\dot{y}$ son también utilizadas.

Es imposible encontrar la derivada haciendo simplemente $x_1 = x$ en la expresión para el cociente incremental, pues entonces tanto el numerador como el denominador serían iguales a cero y esto nos conduciría a la expresión sin significado 0/0. Así, el paso al límite en cada caso depende de ciertos pasos preliminares (la transformación del cociente incremental). Por ejemplo, para la función $f(x) = x^2$ se tiene

$$\frac{f(x_1) - f(x)}{x_1 - x} = \frac{x_1^2 - x^2}{x_1 - x} = x_1 + x \quad \text{siempre que } x \neq x_1.$$

Esta función $x_1 + x$ no posee exactamente el mismo dominio que $(x_1^2 - x^2)/(x_1 - x)$: La función $x_1 + x$ está definida en el punto $x_1 = x$, donde el cociente $(x_1^2 - x^2)/(x_1 - x)$ no está definido. Para cualesquiera otros valores de x_1 las dos funciones son iguales; por lo tanto, al pasar al límite, para el cual se requiere específicamente que $x_1 \neq x$, se obtiene el mismo valor para $\lim\limits_{x_1 \to x} (x_1^2 - x^2)/(x_1 - x)$ que para $\lim\limits_{x_1 \to x} (x_1 + x)$. Sin embargo, puesto que la función $x_1 + x$ está definida y es continua en el punto $x_1 = x$, puede hacerse con ésta lo que no se pudo hacer con el cociente, a saber, pasar al límite haciendo simplemente $x_1 = x$. Para la derivada se obtiene entonces

$$f'(x) = \frac{d(x^2)}{dx} = 2x.$$

Como otro ejemplo, derivamos, esto es, calculamos la derivada de la función $y = \sqrt{x}$ para $x > 0$. Para $x_1 \neq x$ se tiene

$$\frac{f(x_1) - f(x)}{x_1 - x} = \frac{\sqrt{x_1} - \sqrt{x}}{x_1 - x} = \frac{(\sqrt{x_1} - \sqrt{x})(\sqrt{x_1} + \sqrt{x})}{(x_1 - x)(\sqrt{x_1} + \sqrt{x})}$$

$$= \frac{x_1 - x}{(x_1 - x)(\sqrt{x_1} + \sqrt{x})} = \frac{1}{\sqrt{x_1} + \sqrt{x}}.$$

Por lo tanto (para $x > 0$)

$$\frac{d\sqrt{x}}{dx} = \lim_{x_1 \to x} \frac{1}{\sqrt{x_1} + \sqrt{x}} = \frac{1}{2\sqrt{x}}.$$

Para $x = 0$ se tiene una singularidad: La *derivada es infinita,* puesto que $(\sqrt{x_1} - 0)/(x_1 - 0) = 1/\sqrt{x_1} \to \infty$ para $x_1 \to 0$.

Definición analítica

Es sumamente significativo que el proceso de derivar una función posea un significado analítico definido completamente separado de la concepción intuitiva geométrica de la tangente. La definición analítica de la

integral, libre de la visualización geométrica de área, nos permitió basar la noción de área sobre la de integral. Con igual espíritu, independientemente de la representación geométrica de una función $y = f(x)$ por medio de una curva, *definimos* la derivada de la función $y = f(x)$ como la nueva función $y' = f'(x)$ dada mediante el límite del cociente incremental $\Delta y/\Delta x$, siempre que el límite exista.

Aquí las diferencias $\Delta y = y_1 - y = f(x_1) - f(x)$ y $\Delta x = x_1 - x$ son "cambios o variaciones correspondientes" en las variables y y x. La razón $\Delta y/\Delta x$ puede denominarse "razón promedio de variación" de y con respecto a x en el intervalo $(x, x + \Delta x)$. El límite $f'(x) = dy/dx$ representa entonces la "razón instantánea de variación" o simplemente la "razón de variación" de y con respecto a x.

Si este límite existe se dice que la función $f(x)$ es *derivable* o *diferenciable*. Se supondrá siempre que cada función que se considere es derivable a menos que se especifique lo contrario.[1] Se destaca que si la función $f(x)$ ha de ser derivable en el punto x, el límite para $h \to 0$ del cociente $[f(x + h) - f(x)]/h$ debe existir, donde h puede poseer cualquier valor $\neq 0$ para el cual $x + h$ pertenece al dominio de f. Si, en particular, f está definida en todo un intervalo que contiene al punto x en su interior, entonces el límite debe existir *independientemente de la manera en la cual h tiende a* cero, ya sea a través de valores positivos o a través de valores negativos, sin restricción sobre el signo.

Disponiendo ahora de una definición analítica para la derivada $f'(x)$, se toma el ángulo de dirección α con el eje x positivo, dado por la ecuación $\tan \alpha = f'(x)$, como la dirección de la tangente a la curva en el punto (x, y).[2] De este modo, basando la definición geométrica en la analítica, se evitan dificultades que pueden aparecer por la vaguedad de la visualización geométrica. De hecho, se ha *definido* ahora precisamente lo que se entiende por una tangente a la gráfica de $y = f(x)$ en el punto (x, y) y se dispone de un criterio analítico para decidir si una curva posee o no una tangente en un punto dado (x, y).

Funciones monótonas

No obstante, la interpretación objetiva de la derivada como la pendiente de la tangente a la curva es una ayuda altamente provechosa para el entendimiento, aun en discusiones puramente analíticas. Este es el caso de la siguiente proposición basada en la intuición geométrica:

---

[1] Ejemplos en los cuales esta hipótesis no se satisface se darán posteriormente (ver p. 188). Tales ejemplos justifican el mencionar la derivabilidad como una hipótesis si el contexto lo asegura.

[2] El ángulo α no está determinado absolutamente de manera única sino que puede ser substituido por $\alpha \pm \pi$, $\alpha \pm 2\pi$, etc., a menos que se especifique como arriba que $0 \leq \alpha < \pi$.

La función $f(x)$ es monótona creciente cuando $f'(x) > 0$ y monótona decreciente cuando $f'(x) < 0$.

En verdad, si $f'(x)$ es positiva y la curva es recorrida en la dirección de x creciente, entonces la tangente se inclina hacia arriba, esto es, hacia y creciente (α es un "ángulo agudo"); por lo tanto, en el punto en cuestión la curva se eleva conforme x crece; y si, por otra parte, $f'(x)$ es negativa, la tangente se inclina hacia abajo (α es un "ángulo obtuso") y la curva cae conforme x crece (véase Fig. 2.23). Esto será probado analíticamente en las pp. 198, 199.

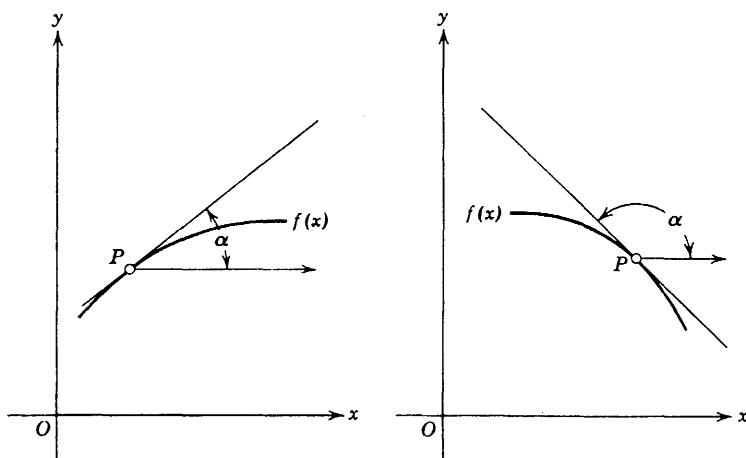


Figura 2.23 Tangentes a gráficas de funciones crecientes y decrecientes.

b. La derivada como una velocidad

La necesidad de substituir el concepto intuitivo de velocidad o rapidez por una definición precisa conduce de nuevo a exactamente el mismo proceso de límite que ya hemos denominado diferenciación.

Considérese el ejemplo de un punto que se mueve sobre una línea recta, el eje y dirigido, estando determinada la posición del punto por una sola coordenada, y. Esta coordenada y es la distancia, con su signo adecuado, de nuestro punto móvil medida desde un punto inicial fijo sobre la línea. El movimiento está dado si se conoce y como función del tiempo t: $y = f(t)$. Si esta función es una función lineal $f(t) = ct + b$, se habla de un *movimiento uniforme con velocidad c*, y para cualquier par de valores distintos, t y t_1, puede obtenerse la velocidad dividiendo la distancia recorrida en un intervalo de tiempo por la longitud de dicho intervalo de tiempo:

$$c = \frac{f(t_1) - f(t)}{t_1 - t}.$$

La velocidad es, por lo tanto, el cociente incremental de la función $ct + b$, y este cociente de incrementos es independiente del par particular de instantes que se fijen. Pero, ¿qué es lo que se ha de entender por la velocidad de movimiento en un instante t si el movimiento no es ya uniforme?

Para responder a esta pregunta, se considera el cociente de incrementos

$$[f(t_1) - f(t)]/(t_1 - t),$$

el cual se denominará *"velocidad promedio"* en el intervalo de tiempo entre t_1 y t. Ahora bien, si esta velocidad promedio tiende a un límite definido cuando se hace tender t_1 a t, este límite se definirá como la velocidad en el tiempo t. En otras palabras: *la velocidad,* esto es, la razón instantánea de variación de la distancia con respecto al tiempo *en el tiempo t, es la derivada*

$$f'(t) = \lim_{t_1 \to t} \frac{f(t_1) - f(t)}{t_1 - t}.$$

Newton destacó la interpretación de derivadas[1] como velocidades, y escribió $\dot{y}$, o bien $\dot{f}(x)$, en vez de $f'(t)$, una notación que se utilizará ocasionalmente. Nuevamente, la derivabilidad de la función es una suposición necesaria si la noción de velocidad ha de tener un significado.

Un ejemplo sencillo es el del movimiento de cuerpos en caída libre. Se parte de la ley establecida experimentalmente según la cual la distancia recorrida en un tiempo t por un cuerpo en caída libre, que parte del reposo en $t = 0$, es proporcional a t^2; es por lo tanto representada mediante una función de la forma

$$y = f(t) = at^2$$

con a constante. Como en las pp. 180-181, la velocidad entonces está dada mediante la expresión $f'(t) = 2at$; así: *la velocidad de un cuerpo en caída libre crece en proporción al tiempo.*

c. Ejemplos de diferenciación

Se ilustrará ahora la técnica de la derivación mediante un número de ejemplos típicos.

Funciones lineales

Para la función $y = f(x) = c$ con c constante, se ve para todo x, es $f(x + h) - f(x) = c - c = 0$, de modo que $\lim_{h \to 0} [f(x + h) - f(x)]/h = 0$; esto es, *la derivada de una función constante es cero.*

---

[1] Denominadas por él "fluxiones".

Para una función lineal $y = f(x) = cx + b$, se encuentra

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{ch}{h} = c.$$

La derivada de una función lineal es constante.

Potencias de x

En seguida se deriva la función potencia

$$y = f(x) = x^{\alpha},$$

suponiendo primero que α es un entero positivo. Siempre que $x_1 \neq x$ se tiene

$$\frac{f(x_1) - f(x)}{x_1 - x} = \frac{x_1^{\alpha} - x^{\alpha}}{x_1 - x} = x_1^{\alpha-1} + x_1^{\alpha-2}x + \cdots + x^{\alpha-1},$$

donde se divide directamente o bien se utiliza la fórmula para la suma de una progresión geométrica. Esta sencilla manipulación algebraica es la clave para pasar al límite; pues la última expresión en el miembro derecho de la ecuación es una función continua de x_1, en particular para $x_1 = x$, y así puede efectuarse el paso al límite con $x_1 \to x$ para esta expresión simplemente reemplazando x_1 en todas partes por x. Cada término toma entonces el valor $x^{\alpha-1}$, y, puesto que el número de términos es exactamente α, se obtiene

$$y' = f'(x) = \frac{d(x^{\alpha})}{dx} = \alpha x^{\alpha-1}.$$

Se llega al mismo resultado si α es un entero negativo $-\beta$; se debe, sin embargo, suponer que x no es cero. Entonces se encuentra

$$\frac{f(x_1) - f(x)}{x_1 - x} = \frac{\frac{1}{x_1^{\beta}} - \frac{1}{x^{\beta}}}{x_1 - x} = -\frac{x^{\beta} - x_1^{\beta}}{x - x_1} \frac{1}{x^{\beta}x_1^{\beta}}$$

$$= -\frac{x^{\beta-1} + x^{\beta-2}x_1 + \cdots + x_1^{\beta-1}}{x_1^{\beta}x^{\beta}}.$$

Una vez más, puede efectuarse el paso al límite simplemente substituyendo x por x_1. Entonces, tal como se hizo antes se obtiene para el límite

$$y' = -\beta \frac{x^{\beta-1}}{x^{2\beta}} = -\beta x^{-\beta-1}.$$

Por lo tanto, para valores enteros *negativos* $\alpha = -\beta$ la derivada está dada nuevamente por la fórmula

$$y' = \alpha x^{\alpha-1}.$$

Finalmente, se probará la misma fórmula cuando x es positivo y α cualquier número racional. Se supone que $\alpha = p/q$, donde p y q son ambos enteros y además positivos. (Si uno de ellos fuera negativo no se requerirían cambios esenciales en la demostración; para $\alpha = 0$ el resultado es ya conocido, puesto que x^α es entonces constante.) Se tiene ahora

$$\frac{f(x_1) - f(x)}{x_1 - x} = \frac{x_1^{p/q} - x^{p/q}}{x_1 - x}.$$

Si se hace ahora $x^{1/q} = \xi$ y $x_1^{1/q} = \xi_1$, se obtiene

$$\frac{f(x_1) - f(x)}{x_1 - x} = \frac{\xi_1{}^p - \xi^p}{\xi_1{}^q - \xi^q} = \frac{\xi_1^{p-1} + \xi_1^{p-2}\xi + \cdots + \xi^{p-1}}{\xi_1^{q-1} + \xi_1^{q-2}\xi + \cdots + \xi^{q-1}}.$$

Después de esta última transformación puede efectuarse inmediatamente el paso al límite con $x_1 \to x$ (o, lo que equivale a lo mismo, $\xi_1 \to \xi$) y obtener así para el valor límite la expresión

$$y = \frac{p}{q}\frac{\xi^{p-1}}{\xi^{q-1}} = \frac{p}{q}\xi^{p-q} = \frac{p}{q}x^{(p-q)/q} = \frac{p}{q}x^{(p/q)-1}$$

o, finalmente,

$$f'(x) = y' = \alpha x^{\alpha-1},$$

que es formalmente el mismo resultado obtenido antes. Se deja al lector probar que la misma fórmula de derivación vale también para exponentes racionales negativos.

Se volverá (p. 207) a la derivación de potencias y se probará la validez general de la fórmula anterior para exponentes arbitrarios α.

Funciones trigonométricas

Como un último ejemplo se considera la *derivación de las funciones trigonométricas* sen x y cos x. Se utiliza la elemental fórmula de adición trigonométrica para transformar el cociente incremental

$$\frac{\operatorname{sen}(x+h) - \operatorname{sen} x}{h} = \frac{\operatorname{sen} x \cos h + \cos x \operatorname{sen} h - \operatorname{sen} x}{h}$$

$$= \operatorname{sen} x \frac{\cos h - 1}{h} + \cos x \frac{\operatorname{sen} h}{h}.$$

Recordando las relaciones de la sección 1.8, pp. 107-108,

$$\lim_{h\to 0} \frac{\operatorname{sen} h}{h} = 1, \qquad \lim_{h\to 0} \frac{\cos h - 1}{h} = 0,$$

y se obtiene inmediatamente

$$y' = \frac{d(\operatorname{sen} x)}{dx} = \cos x.$$

La función $y = \cos x$ puede ser derivada exactamente de la misma forma. Partiendo con

$$\frac{\cos(x+h) - \cos x}{h} = \cos x \frac{\cos h - 1}{h} - \operatorname{sen} x \frac{\operatorname{sen} h}{h}$$

y tomando el límite para $h \to 0$, se obtiene la derivada [1]

$$y' = \frac{d(\cos x)}{dx} = -\operatorname{sen} x.$$

d. Algunas reglas fundamentales para la derivación

Tal como en el caso de la integral, existen ciertas reglas básicas para la derivación que se obtienen inmediatamente a partir de la definición y bastan para formar la derivada de muchas funciones.

1. Si $\phi(x) = f(x) + g(x)$, entonces $\phi'(x) = f'(x) + g'(x)$.
2. Si $\psi(x) = cf(x)$ (donde c es una constante), entonces $\psi'(x) = cf'(x)$. Se tiene

$$\frac{\phi(x+h) - \phi(x)}{h} = \frac{f(x+h) - f(x)}{h} + \frac{g(x+h) - g(x)}{h}$$

y

$$\frac{\psi(x+h) - \psi(x)}{h} = c\frac{f(x+h) - f(x)}{h},$$

y nuestras proposiciones se obtienen directamente mediante el paso al límite.

Así, por ejemplo, la derivada de la función $\phi(x) = f(x) + ax + b$ (donde a y b son constantes) está dada por la ecuación

$$\phi'(x) = f'(x) + a.$$

Con la ayuda de estas reglas y de la fórmula para la derivada de una potencia se puede inmediatamente derivar cualquier polinomio $y = a_0x^n + a_1x^{n-1} + \cdots + a_n$ y encontrar

$$y' = na_0x^{n-1} + (n-1)a_1x^{n-2} + \cdots + 2a_{n-2}x + a_{n-1}.$$

e. Derivabilidad y continuidad de funciones

Es útil saber que la derivabilidad es una condición más fuerte que la continuidad:

---

[1] Si x es interpretado como un ángulo, entonces estas fórmulas sencillas para las derivadas de $\operatorname{sen} x$ y $\cos x$ presuponen, por supuesto, que el ángulo x está medido en radianes.

Si una función es derivable, entonces es también continua.

Pues si el cociente de incrementos $[f(x+h) - f(x)]/h$ se aproxima a un límite definido cuando h tiende a cero, el numerador de la fracción, esto es, $f(x+h) - f(x)$ debe[1] tender a cero con h; esto precisamente expresa la continuidad de la función $f(x)$ en el punto x. Por lo tanto, son innecesarias otras demostraciones engorrosas de la continuidad de funciones que puede mostrarse que son derivables (esto es, para la mayoría de las funciones que se encontrarán).

Discontinuidades de la derivada-Esquinas

La recíproca, sin embargo, es falsa; *no* es cierto que toda función continua posea una derivada en cada punto. El contraejemplo más sencillo es la función $f(x) = |x|$, esto es, $f(x) = -x$ para $x \leq 0$ y $f(x) = x$ para $x \geq 0$; su gráfica está mostrada en la Fig. 2.24. En el punto $x = 0$ esta

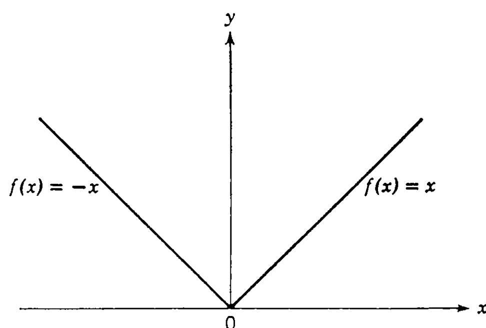


Figura 2.24 $f(x) = |x|$.

función es continua pero no posee derivada. El límite de $[f(x+h) - f(x)]/h$ es igual a 1 si h tiende a cero a través de valores positivos, y es igual a -1 si h tiende a cero por valores negativos; y si el signo de h no se restringe, ningún límite existe. Se dice que nuestra función posee diferentes *derivadas anterior* y *posterior* en el punto $x = 0$, donde por derivada anterior y derivada posterior se entienden respectivamente los valores límite de $[f(x+h) - f(x)]/h$ conforme h se aproxima a cero sólo a través de valores positivos y sólo a través de valores negativos respectivamente. La *derivabilidad* de una función definida en un intervalo en torno al punto considerado requiere así no solamente que existan las derivadas anterior y posterior, sino que éstas sean iguales. Geométricamente, la desigualdad

---

[1] Puesto que entonces

$$\lim_{h\to 0} [f(x+h) - f(x)] = \left[\lim_{h\to 0} \frac{f(x+h) - f(x)}{h}\right] (\lim_{h\to 0} h) = f'(x) \cdot 0 = 0.$$

de las dos derivadas significa que la curva posee una *esquina*. La derivabilidad expresa de una manera precisa lo que intuitivamente podría denominarse *suavidad* de la gráfica de la función.

Discontinuidades infinitas

Como un ejemplo más de puntos donde una función continua no es derivable, se consideran los puntos donde la derivada se torna infinita, esto es, los puntos en los cuales ni existe una derivada anterior ni una derivada posterior, y el cociente incremental $[f(x+h) - f(x)]/h$ crece más allá de toda cota conforme $h \to 0$. Por ejemplo, la función $y = f(x) = \sqrt[3]{x} = x^{1/3}$ está definida y es continua para todos los valores de x. Para todos los valores de x distintos de cero su derivada está dada (p. 185) por la fórmula $y' = \frac{1}{3}x^{-2/3}$. En el punto $x = 0$ se tiene $[f(x+h) - f(x)]/h = h^{1/3}/h = h^{-2/3}$, y se ve de inmediato que cuando $h \to 0$ la expresión no posee valor límite sino que, por el contrario, tiende a infinito. Este estado de cosas es a menudo descrito brevemente diciendo que la función posee una derivada infinita en el punto en cuestión; como se podrá recordar, sin embargo, esto significa solamente que conforme h tiende a cero el cociente diferencia crece más allá de toda cota y que la derivada en el sentido en que se ha definido realmente *no existe*. El significado geométrico de una *derivada infinita* es que la tangente a la curva es vertical (ver Fig. 2.25).

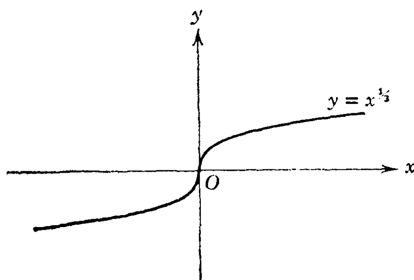


Figura 2.25.

La función $y = f(x) = \sqrt{x}$, la cual está definida y es continua para $x \geq 0$, tampoco es derivable en el punto $x = 0$. Puesto que y no está definida para valores negativos de x, se considera aquí solamente la derivada lateral derecha. La ecuación $[f(h) - f(0)]/h = 1/\sqrt{h}$ muestra que esta derivada es infinita; la curva toca al eje y en el origen (Fig. 2.26).

Finalmente, en la función $y = \sqrt[3]{x^2} = x^{2/3}$ se tiene un caso en el cual la derivada lateral derecha en el punto $x = 0$ es positiva e infinita, en tanto que la derivada lateral izquierda es negativa e infinita, como se sigue de la relación

$$\frac{f(h) - f(0)}{h} = \frac{1}{\sqrt[3]{h}}$$

En realidad, la curva continua $y = x^{2/3}$, llamada *parábola semicúbica* o *parábola de Neil,* posee en el origen una *cúspide* con una tangente perpendicular al eje x (ver Fig. 2.27).

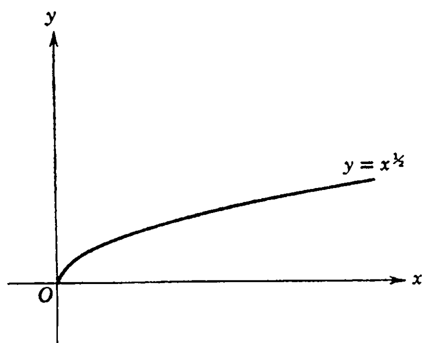


Figura 2.26.

f. Derivadas superiores y su significado

La gráfica de la derivada $f'(x)$ de una función se denomina la *curva derivada* de la gráfica de $f(x)$. Por ejemplo, la curva derivada de la parábola $y = x^2$ es una línea recta representada por la función $y = 2x$. La curva derivada de la curva seno $y = \operatorname{sen} x$ es la curva coseno $y = \cos x$; y, análogamente, la curva derivada de la curva $y = \cos x$ es la curva $y = -\operatorname{sen} x$. (Estas últimas curvas pueden obtenerse una de la otra mediante

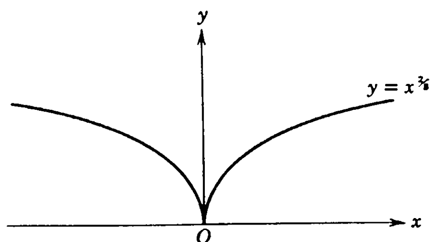


Figura 2.27.

una translación en la dirección del eje x, como se muestra en la Fig. 2.28.)

Es bastante natural el formar curvas derivadas de las curvas derivadas, esto es, formar la derivada de la función $f'(x) = \phi(x)$. Esta derivada

$$\phi'(x) = \lim_{h \to 0} \frac{f'(x + h) - f'(x)}{h}$$

suponiendo que existe, se denomina *segunda derivada* de la función $f(x)$; y se la denotará por $f''(x)$.

Análogamente, puede intentarse formar la derivada de $f''(x)$, llamada *tercera derivada* de $f(x)$, la cual entonces se denota por $f'''(x)$. Para la mayoría de las funciones que nos conciernen no hay nada que nos impida repetir el proceso de derivación tantas veces como queramos, definiendo así una n-ésima *derivada* $f^{(n)}(x)$.[1] Ocasionalmente, será conveniente denominar a la función $f(x)$ su propia 0-ésima derivada.

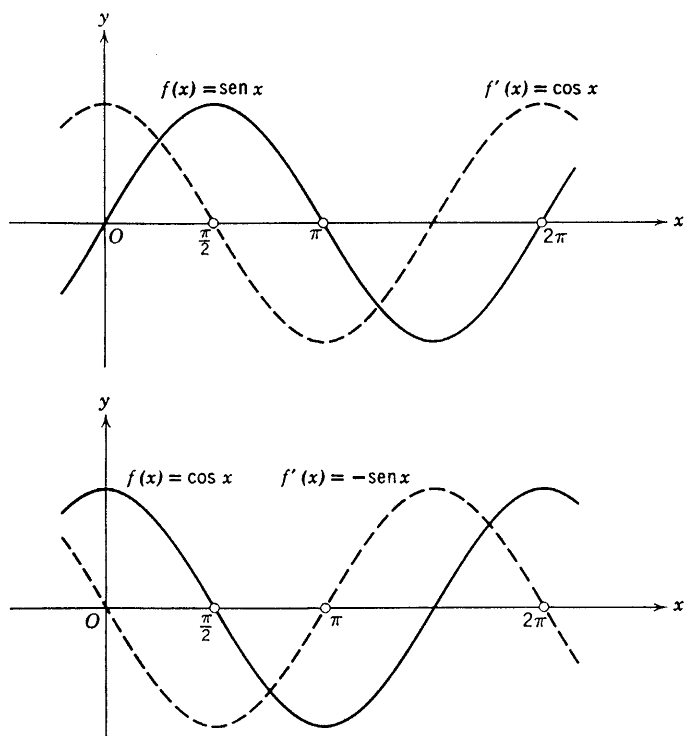


Figura 2.28 Curvas derivadas de sen x y cos x.

Si la variable independiente es interpretada como el tiempo t y el movimiento de un punto está representado como antes por la función $f(t)$, el significado *físico* de la segunda derivada es la razón de variación de la velocidad $f'(t)$ con respecto al tiempo, o sea, como se le denomina usualmente, la *aceleración*. En el ejemplo del cuerpo en caída libre la distancia recorrida en el tiempo t estaba dada mediante la función $y = f(t) = at^2$. Se encontró $f'(t) = 2at$ para la velocidad en el tiempo t. La aceleración

[1] Los términos *segundo, tercero, ... , n*-ésimo *coeficiente diferencial* son también utilizados, o bien $D^2f, \ldots, D^nf$ (ver nota 3, p. 180).

posee entonces el valor constante $f''(t) = 2a$ (el cual usualmente es identificado con la constante gravitacional g). Posteriormente (p. 257), se discutirá la interpretación geométrica de la segunda derivada en detalle. Aquí, sin embargo, se toma nota de los siguientes hechos: En un punto donde $f''(x)$ es positiva, $f'(x)$ crece conforme x crece; si aquí $f'(x)$ es positiva, la curva se torna más empinada para x creciente. Si por otra parte, $f''(x)$ es negativa, $f'(x)$ decrece conforme x crece, y si $f'(x)$ es positiva, la curva se torna menos empinada conforme x crece.

Finalmente, se observa que las derivadas superiores pueden utilizarse para definir una función. Así, las funciones trigonométricas se pueden caracterizar por una así llamada *ecuación diferencial* que involucra la función y su segunda derivada. De las fórmulas $(d \cos x)/dx = -\operatorname{sen} x$, $(d \operatorname{sen} x)/dx = \cos x$ se obtiene inmediatamente al derivar nuevamente,

$$\frac{d^2}{dx^2}\cos x = -\cos x, \qquad \frac{d^2}{dx^2}\operatorname{sen} x = -\operatorname{sen} x.$$

Por tanto, si el símbolo u es representativo de cualquiera de las funciones $\operatorname{sen} x$ o $\cos x$, se tiene la relación *(ecuación diferencial)*

$$u'' = -u.$$

Esta ecuación diferencial claramente la satisface también cualquier combinación lineal $u = a \cos x + b \operatorname{sen} x$ con coeficientes constantes a, b. Se verá en la Sec. 3.16 b que tales combinaciones lineales, con constantes arbitrarias a y b, son las únicas funciones u para las cuales $u'' = -u$.

En todos los tipos de aplicaciones que involucran oscilaciones o fenómenos ondulatorios, tales como movimientos de resortes o bien ondas sobre la superficie del agua, pasaremos directamente de consideraciones físicas hacia una ecuación diferencial del tipo $u'' = -u$ para la variable física significativa u (usualmente la variable independiente es el *tiempo*). Es por ello importante reconocer que u puede ser representada sencillamente en términos de funciones trigonométricas (véase capítulo 9).

g. Derivada y cociente de incrementos. Notación de Leibnitz

En la notación de Leibnitz el paso al límite en el proceso de derivación es simbólicamente expresado reemplazando el símbolo Δ por el símbolo d, motivando el símbolo de Leibnitz para la derivada *definida* mediante la ecuación

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}.$$

Si se desea obtener un entendimiento claro del significado del cálculo diferencial, debemos precavernos contra la vieja falacia de imaginar la derivada como el cociente de dos "cantidades" dy y dx que son de hecho

"infinitamente pequeñas". El *cociente de incrementos* $\Delta y/\Delta x$ posee un significado sólo para diferencias Δx que no son iguales a cero. *Después* de formar este genuino cociente incremental debe efectuarse el paso al límite por medio de una transformación o algún otro plan el cual también en el límite evita la división por cero. No tiene sentido suponer que *primero* Δx y Δy atraviesan por algo parecido a un proceso de límite y alcanzan valores que son infinitesimalmente pequeños pero aún diferentes de cero, de modo que Δx y Δy se reemplazan por "cantidades infinitamente pequeñas" o "infinitesimales" dx y dy, y que el cociente de estas cantidades es entonces formado. Tal concepción de la derivada es incompatible con la claridad matemática; de hecho, no tiene significado alguno. Para mucha gente indudablemente posee un cierto encanto de misterio, siempre asociado con la palabra "infinito"; y en los primeros días del cálculo diferencial aun el mismo Leibnitz fue capaz de combinar estas vagas ideas místicas con un tratamiento enteramente claro de los procesos de límite. Pero ahora el misticismo de las cantidades "infinitamente pequeñas" no tiene lugar en el cálculo.

La notación de Leibnitz, sin embargo, no es meramente sugestiva por sí misma, sino que es en realidad sumamente flexible y útil. La razón es que en muchos cálculos y transformaciones formales podemos tratar con los símbolos dy y dx exactamente *como si* fueran números ordinarios. Tratando a dx y dy como números es posible dar una expresión más clara a muchos cálculos que pueden llevarse a cabo con toda validez sin su uso. En los siguientes capítulos se verá este hecho verificado una y otra vez, y nos habremos justificado al hacer un uso libre y repetido de ellos, siempre que no se pierda de vista el carácter simbólico de los signos dy y dx.

*También para las derivadas segunda y superiores Leibnitz ideó una notación sugestiva. Consideró la segunda derivada como el límite del "segundo cociente de incrementos" de la siguiente manera: Además de la variable x se considera $x_1 = x + h$ y $x_2 = x + 2h$. Entonces se toma el segundo cociente de incrementos, entendiéndose por éste el primer cociente de incrementos del primer cociente, esto es, la expresión

$$\frac{1}{h}\left(\frac{y_2 - y_1}{h} - \frac{y_1 - y}{h}\right) = \frac{1}{h^2}(y_2 - 2y_1 + y),$$

donde $y = f(x)$, $y_1 = f(x_1)$ y $y_2 = f(x_2)$. Escribiendo $h = \Delta x, y_2 - y_1 = \Delta y_1$ y $y_1 - y = \Delta y$, se puede denominar apropiadamente la expresión en el último paréntesis "diferencia de la diferencia" de y, o bien, *segunda diferencia* de y, y escribir simbólicamente [1]

$$y_2 - 2y_1 + y = \Delta y_1 - \Delta y = \Delta(\Delta y) = \Delta^2 y.$$

[1] Aquí $\Delta\Delta = \Delta^2$ es simplemente un símbolo para "diferencia de diferencia" o bien "segunda diferencia".

En esta notación simbólica el segundo cociente incremental es escrito entonces $\Delta^2 y/(\Delta x)^2$, donde el denominador es realmente el cuadrado de Δx, mientras que en el numerador el índice superior 2 denota simbólicamente la repetición del proceso de diferencia. La segunda derivada es expresada entonces mediante

$$f''(x) = \lim_{x_1 \to x} \frac{\Delta^2 f}{(\Delta x)^2}.$$

Este simbolismo para el cociente incremental[1] condujo a Leibnitz a introducir la notación

$$y'' = f''(x) = \frac{d^2y}{dx^2}, \qquad y''' = f'''(x) = \frac{d^3y}{dx^3}, \quad \text{etc.},$$

para las derivadas segundas y superiores, y se verá que esta notación es también de utilidad.[2]

h. El teorema del valor medio del cálculo diferencial

El cociente incremental involucra los valores de una función para distintos valores de x, mientras que la derivada en un punto no nos dice nada acerca de la función en cualquier otro punto; el cociente incremental refleja propiedades de la función "globalmente", mientras la derivada refleja una propiedad *local,* o sea, una propiedad "en lo pequeño". A menudo se requerirá deducir propiedades del todo, o propiedades "globales", de una función a partir de las propiedades locales dadas por su derivada. Para este propósito se utiliza una relación fundamental entre el cociente de incrementos y la derivada, conocido como "el teorema del valor medio del cálculo diferencial".

El teorema del valor medio es fácil de apreciar intuitivamente. Considérese el cociente incremental

$$\frac{f(x_1) - f(x_2)}{x_1 - x_2} = \frac{\Delta f}{\Delta x}$$

de una función $f(x)$ y supóngase que la derivada existe en todas partes en el intervalo cerrado $x_1 \leq x \leq x_2$, de modo que la gráfica de la curva posea una tangente en todas partes. El cociente de incrementos es la tan-

---

[1] Como debe destacarse, la proposición de que la segunda derivada puede ser representada como el límite del segundo cociente de incrementos requiere, una demostración. Previamente, la segunda derivada no se definió de esta forma sino como el límite del primer cociente de incrementos de la primera derivada. Las dos definiciones son equivalentes siempre que la segunda derivada sea *continua;* la demostración, sin embargo, se dará sólo posteriormente (véase el capítulo 5, apéndice II), puesto que no tenemos aún ninguna necesidad particular de esta propiedad.

[2] Esta es la notación que se acostumbra. Escribiendo $y'' = d^2y/(dx)^2$, $y''' = d^3y/(dx)^3$, con paréntesis, sería un poco más clara, pero de ordinario no se hace.

gente del ángulo de inclinación α de la secante, mostrada en la Fig. 2.29. Imagínese esta secante trasladada paralelamente a sí misma. Al menos una vez alcanzará una posición en la cual es una tangente a la curva en un punto entre x_1 y x_2; esto ocurre en un punto $x = \xi$ de la curva el cual se encuentra a la máxima distancia de la secante. Por lo tanto existe un valor intermedio ξ en el intervalo, tal que

$$\frac{f(x_1) - f(x_2)}{x_1 - x_2} = f'(\xi).$$

Esta proposición se denomina *teorema del valor medio del cálculo diferencial.*[1] Puede expresarse también en forma algo diferente observando que el número ξ puede ser escrito en la forma

$$\xi = x_1 + \Theta(x_2 - x_1),$$

donde todo lo que se sabe en torno a Θ es que se encuentra situado entre 0 y 1. Aún cuando Θ (o bien ξ) en general no puede ser especificado más exactamente, el teorema es sumamente poderoso en las aplicaciones.

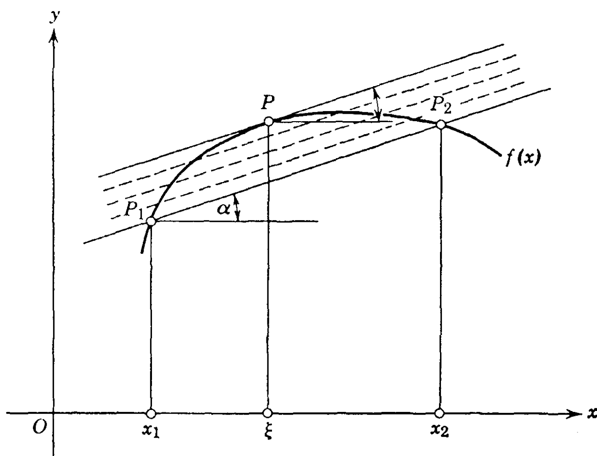


Figura 2.29.

Considérese, por ejemplo, el caso donde x es el tiempo y $y = f(x)$ la distancia de un automóvil desde su punto de partida a lo largo de cierto camino. Entonces $f'(x)$ es la velocidad del automóvil en el tiempo x. Si, digamos, durante las dos primeras horas ($\Delta x = 2$) el automóvil ha cubierto una distancia $\Delta f = 120$ millas, puede concluirse, a partir del teorema del valor medio, que al menos en un momento ξ durante estas dos horas el automóvil tuvo una velocidad de exactamente 60 millas por hora (siem-

[1] Un nombre más apropiado sería el de *teorema del valor intermedio* del cálculo diferencial.

pre que la velocidad exista en todo momento). El conductor no puede sostener, por ejemplo, que ha viajado todo el tiempo a menos de 50 millas por hora. Por otra parte, no hay nada que indique cuál fue el tiempo ξ en el cual se alcanzó la velocidad precisa de 60 millas por hora; pudo haber sido en algún tiempo durante la primera hora, o bien durante la segunda hora o bien en diversas ocasiones.

Una formulación precisa del teorema del valor medio es la siguiente:

Si $f(x)$ es continua en el intervalo cerrado $x_1 \leq x \leq x_2$ y derivable en todo punto del intervalo abierto $x_1 < x < x_2$, entonces existe al menos un valor θ, con $0 < \theta < 1$, tal que

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = f'[x_1 + \theta(x_2 - x_1)].$$

Si se substituye x_1 por x y x_2 por $x + h$, el teorema del valor medio puede ser expresado mediante la fórmula

$$\frac{f(x + h) - f(x)}{h} = f'(\xi) = f'(x + \theta h), \qquad x < \xi < x + h.$$

Aun cuando es esencial que $f(x)$ sea continua para todos los puntos del intervalo, incluyendo los puntos extremos, no se tiene que suponer que la derivada exista en los puntos extremos.

Si en cualquier punto en el interior del intervalo la derivada deja de existir, el teorema del valor medio no es necesariamente cierto. Es fácil ver esto en el ejemplo de $f(x) = |x|$.

i. Demostración del teorema

El teorema del valor medio se deduce usualmente mediante la reducción a un caso especial que será establecido primero.

TEOREMA DE ROLLE. *Si una función $\phi(x)$ es continua en el intervalo cerrado $x_1 \leq x \leq x_2$ y derivable en el intervalo abierto $x_1 < x < x_2$, y si, además, $\phi(x_1) = 0$ y $\phi(x_2) = 0$, entonces existe al menos un punto ξ en el interior del intervalo en el cual $\phi'(\xi) = 0$.*

Interpretado geométricamente, esto significa que si una curva alcanza el eje x en dos puntos, entonces debe poseer una tangente horizontal en algún punto intermedio (Fig. 2.30).

Sin duda, puesto que $\phi(x)$ es continua en el intervalo cerrado $[x_1, x_2]$, existe un valor máximo M de $\phi(x)$ y un valor mínimo m en este intervalo (véase p. 124). Puesto que ϕ se hace cero en los puntos extremos, se debe tener $m \leq 0 \leq M$. Si estos valores máximo y mínimo deben ser iguales, entonces necesariamente $m = M = 0$ y así $\phi(x) = 0$ en todos los puntos

del intervalo; entonces también $\phi'(x) = 0$ en el intervalo, y, por lo tanto, $\phi'(\xi) = 0$ para todo ξ en el intervalo. Entonces, debe considerarse solamente el caso donde m y M no son cero ambos. Si, en particular M no es cero, entonces M debe ser positivo. Existe un punto en el intervalo $[x_1, x_2]$ donde $\phi(\xi) = M$. Puesto que ϕ se hace cero en los puntos extremos del intervalo, el punto ξ debe ser un punto interior. Además, $\phi(x) \leq \phi(\xi) = M$

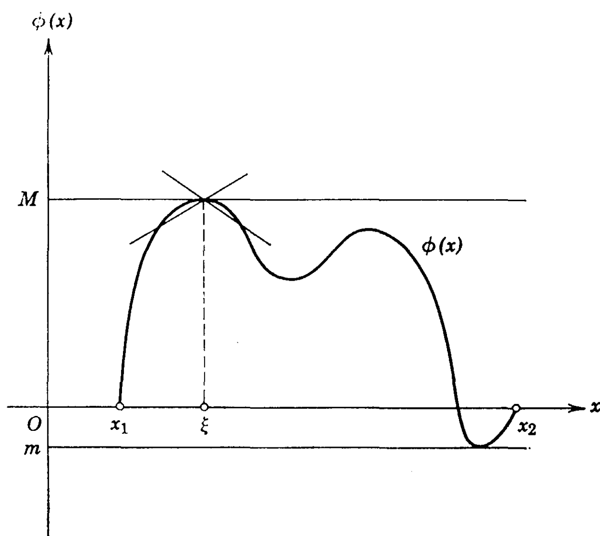


Figura 2.30.

para todo x en $[x_1, x_2]$. Consecuentemente, para todo número h cuyo valor absoluto $|h|$ es suficientemente pequeño, la desigualdad $\phi(\xi + h) - \phi(\xi) \leq 0$ se cumple. Esto implica que el cociente

$$\frac{\phi(\xi + h) - \phi(\xi)}{h}$$

es negativo o bien cero para $h > 0$ y positivo o bien cero para $h < 0$. Si se hace h tender a cero a través de valores positivos se encuentra que $\phi'(\xi) \leq 0$, mientras que para h tendiendo a cero a través de valores negativos se sigue que $\phi'(\xi) \geq 0$. Por lo tanto, $\phi'(\xi) = 0$; y así el teorema de Rolle ha sido probado en el caso $M \neq 0$. El mismo argumento se cumple para $m \neq 0$.

Para probar el teorema del valor medio se aplica el teorema de Rolle a una función que representa la distancia vertical entre el punto $(x, f(x))$ de la gráfica y su secante:

$$\phi(x) = f(x) - f(x_1) - \frac{f(x_2) - f(x_1)}{x_2 - x_1}(x - x_1).$$

Esta función[1] obviamente satisface la condición $\phi(x_1) = \phi(x_2) = 0$, y es de la forma $\phi(x) = f(x) + ax + b$ con coeficientes constantes $a = -[f(x_2) - f(x_1)]/(x_2 - x_1)$ y b. De la p. 187 se sabe que

$$\phi'(x) = f'(x) + a,$$

y así, por el teorema de Rolle,

$$0 = \phi'(\xi) = f'(\xi) + a$$

para un valor intermedio ξ elegido adecuadamente; por lo tanto,

$$f'(\xi) = -a = \frac{f(x_2) - f(x_1)}{x_2 - x_1};$$

y así el teorema del valor medio está probado.

Significado del teorema

La derivada de una función fue definida como el límite de cocientes de incrementos para un intervalo cuando los puntos extremos se aproximan entre sí. El teorema del valor medio establece una conexión entre cocientes de incrementos y derivadas de una función derivable, que no involucra la contracción del intervalo. Cada cociente incremental es igual a la derivada en un punto intermedio adecuado ξ.

Ejemplos. Tal como en el teorema del valor medio del cálculo integral, no hay nada afirmado específicamente en el teorema del valor intermedio respecto a la ubicación de ξ además del hecho de que ξ está situado en el interior del intervalo. Para el ejemplo de la función cuadrática $y = f(x) = x^2$ con derivada $f'(x) = 2x$, se encuentra

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = x_1 + x_2 = f'(\xi),$$

donde $\xi = (x_1 + x_2)/2$ es el punto medio del intervalo $[x_1, x_2]$. En general, sin embargo, ξ puede estar situado en cualquier otra parte entre x_1 y x_2. Por ejemplo, si $f(x) = x^3$, se tiene $[f(1) - f(0)]/(1 - 0) = 1 = f'(\xi) = 3\xi^2$, donde $\xi = 1/\sqrt{3}$.

Funciones monótonas. Como una de las muchas aplicaciones del teorema del valor medio del cálculo diferencial, se prueba que si la derivada de

[1] Esta función es también proporcional a la distancia del punto $(x, f(x))$, de la curva, a la secante. El lector puede verificar esto por sí mismo fácilmente; por ejemplo, utilizando la propiedad de la geometría analítica elemental de que la expresión $(y - mx - b)/\sqrt{1 + m^2}$ representa la distancia (con su signo) del punto (x, y) a la recta con ecuación $y - mx - b = 0$. De esta forma se encuentra que, efectivamente, en los puntos de la curva que están a una distancia mayor de la secante, la tangente es paralela a la secante.

$f(x)$ posee un signo constante, entonces f es monótona. Específicamente, se supone $f(x)$ continua en el intervalo cerrado $[a, b]$ y derivable en cada punto del intervalo abierto (a, b). *Si en estas condiciones es* $f'(x) > 0$ *para* x *en* (a, b), *entonces la función* $f(x)$ *es monótona creciente; similarmente, si* $f'(x) < 0$, *la función es monótona decreciente.* La demostración es obvia: Sean x_1 y x_2 dos valores cualesquiera en el intervalo cerrado $[a, b]$. Entonces existe un ξ entre x_1 y x_2, y, por lo tanto, también entre a y b, tal que

$$f(x_2) - f(x_1) = f'(\xi)(x_2 - x_1).$$

Si $f'(x) > 0$ en todas partes en (a, b), se tiene en particular $f'(\xi) > 0$. Por lo tanto, $f(x_2) - f(x_1)$ es positivo para $x_2 > x_1$; esto es, $f(x)$ es creciente. Análogamente, f es decreciente si $f'(x) < 0$ en (a, b).

De la misma manera se muestra que *una función* $f(x)$ *continua en* $[a, b]$ *y derivable en el intervalo abierto* (a, b) *debe ser constante si* $f'(x) = 0$ *en todas partes en* (a, b). Pues entonces

$$f(x_2) - f(x_1) = f'(\xi)(x_2 - x_1) = 0.$$

Esta importante propiedad corresponde al hecho intuitivamente obvio de que una curva cuya tangente en cada punto es paralela al eje x debe ser una línea recta paralela al eje x.

Continuidad de Lipschitz de funciones diferenciables. Se mencionó anteriormente que una función $f(x)$ que posee una derivada es necesariamente continua. El teorema del valor medio del cálculo diferencial proporciona una información cuantitativa mucho más precisa, a saber, un *módulo de continuidad.* Considérese una función $f(x)$ la cual está definida en el intervalo cerrado $[a, b]$ y posee una derivada $f'(x)$ en cada punto de ese intervalo. Supóngase que $f'(x)$ es *acotada* en el intervalo (éste es ciertamente el caso siempre que $f'(x)$ sea definida y continua en el intervalo cerrado $[a, b]$); existe un número M tal que $|f'(x)| \leq M$. Para dos valores cualesquiera x_1, x_2 en (a, b), se infiere del teorema del valor medio que

$$|f(x_2) - f(x_1)| = |f'(\xi)(x_2 - x_1)| \leq M|x_2 - x_1|.$$

Para $\varepsilon > 0$ dado hemos así producido un módulo sencillo de continuidad, $\delta = \varepsilon/M$, tal que

$$|f(x_2) - f(x_1)| \leq \varepsilon \qquad \text{para} \quad |x_2 - x_1| \leq \delta.$$

Tómese, por ejemplo, la función $f(x) = x^2$ en el intervalo $-a \leq x \leq +a$. Puesto que

$$|f'(x)| = |2x| \leq 2a,$$

se ve que aquí

$$|f(x_2) - f(x_1)| \leq \varepsilon \qquad \text{para} \quad |x_2 - x_1| \leq \varepsilon/2a.$$

Se dijo que una función $f(x)$ "satisface una condición de Lipschitz" o bien que es "continua en el sentido de Lipschitz" (también, que es "lipschitziana") si existe una constante M tal que

$$|f(x_2) - f(x_1)| \leq M\,|x_2 - x_1|$$

para todos x_1, x_2 en cuestión. Esto significa que todos los *cocientes de incrementos*

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

poseen la misma cota superior M para su valor absoluto. Se ve que toda función f con derivada continua f' en un intervalo cerrado es continua en el sentido de Lipschitz. Sin embargo, aun funciones que no poseen derivada en cada punto pueden ser "continuas de Lipschitz", como lo muestra el ejemplo $f(x) = |x|$. El lector puede verificar por él mismo que para esta función siempre es $|f(x_2) - f(x_1)| \leq |x_2 - x_1|$.

Por otra parte, no toda función es continua en el sentido de Lipschitz. Esto se muestra mediante el ejemplo de $f(x) = x^{1/3}$; aquí el cociente

$$\frac{f(x) - f(0)}{x - 0} = x^{-2/3}$$

no está acotado para x pequeño; por lo tanto, $f(x)$ no es continua en el sentido de Lipschitz en $x = 0$. Esto es compatible con el hecho de que la derivada $f'(x) = 1/3x^{2/3}$ no permanece acotada cuando x tiende a cero. Las funciones que son continuas en el sentido de Lipschitz constituyen una clase importante intermedia entre aquéllas que son simplemente continuas y aquéllas que poseen una derivada continua.

j. La aproximación de funciones mediante funciones lineales. Definición de diferenciales

Definición. La derivada de una función $y = f(x)$ se definió mediante

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x},$$

donde $\Delta x = h$. Si para un x fijo y una h variable se define una cantidad ε mediante

$$\varepsilon(h) = \frac{f(x+h) - f(x)}{h} - f'(x) = \frac{\Delta y}{\Delta x} - f'(x),$$

entonces el hecho de que $f'(x)$ es la derivada de f en el punto x equivale a la ecuación

$$\lim_{h \to 0} \varepsilon(h) = 0.$$

La cantidad $\Delta y = f(x + h) - f(x)$ representa la variación o *incremento* en el valor de la variable dependiente y que resulta cuando el valor x de la variable independiente es variado en la cantidad $\Delta x = h$. Puesto que

$$\Delta y = f'(x)\,\Delta x + \varepsilon\,\Delta x,$$

la cantidad Δy aparece como la suma de dos partes, a saber, una parte $f'(x)\,\Delta x$ que es proporcional a Δx y una parte $\varepsilon\,\Delta x$ que puede hacerse tan pequeña como se desee comparada con Δx haciendo a Δx mismo suficientemente pequeño. La parte lineal, dominante, en la expresión para Δy la denominaremos la *diferencial* dy de y, y escribimos

$$dy = df(x) = f'(x)\,\Delta x.$$

Para cualquier función derivable f y para un x fijo esta diferencial es una función lineal bien definida de $h = \Delta x$. Por ejemplo, para la función $y = x^2$ se tiene $dy = d(x^2) = 2x\,\Delta x = 2xh$. Para la función particular $y = x$, cuya derivada posee el valor constante uno, se tiene simplemente $dx = \Delta x$. Es entonces compatible con nuestra definición el escribir dx para Δx cuando x es la variable independiente; por lo tanto, la diferencial de cualquier función $y = f(x)$ puede ser escrita también como

$$dy = df(x) = f'(x)\,dx.$$

El incremento de la variable dependiente

$$\Delta y = f'(x)\,dx + \varepsilon\,dx = dy + \varepsilon\,dx$$

difiere de la diferencial dy en la cantidad $\varepsilon\,dx$, que, en general, no es cero. En el ejemplo de la función $y = x^2$ se tiene $dy = 2xdx$, mientras que

$$\Delta y = (x + dx)^2 - x^2 = 2x\,dx + (dx)^2 = dy + \varepsilon\,dx,$$

donde $\varepsilon = dx$.

Anteriormente se utilizó el símbolo dy/dx de manera puramente simbólica para denotar el límite del cociente $\Delta y/\Delta x$ para Δx tendiendo a cero. Con la presente definición de las diferenciales dy y dx, la derivada dy/dx puede ser considerada en realidad como el cociente ordinario de dy y dx. Aquí, sin embargo, dy y dx no son ahora en ningún sentido cantidades "infinitamente pequeñas" o "infinitesimales"; tal interpretación estaría desprovista de sentido. En vez de eso, dy y dx son funciones lineales bien definidas de $h = \Delta x$, las cuales para Δx grande pueden poseer valores numéricos grandes. No hay nada sobresaliente en el hecho de que el cociente dy/dx de estas cantidades posea el mismo valor que la derivada $f'(x)$. Esta es solamente una tautología reestableciendo la definición de dy como $f'(x)\,dx$.[1]

---

[1] Análogamente, las diferenciales de orden superior podrían ser definidas por $d^2y = f''(x)h^2 = f''(x)(dx)^2$, $d^3y = f'''(x)(dx)^3$, etc., en concordancia con la notación de Leibnitz para las derivadas superiores.

Reescribiendo la relación entre incremento y diferencial de f en la forma

$$f(x+h) = f(x) + hf'(x) + \varepsilon h,$$

se ve que la expresión $f(x+h)$ considerada como una función de h está representada por la función lineal $f(x) + hf'(x)$, con un error εh que es arbitrariamente pequeño comparado con h si h es suficientemente pequeño. Esta representación aproximada de $f(x+h)$ mediante la función lineal $f(x) + hf'(x)$ significa geométricamente que se reemplaza la curva por su tangente en el punto x (véase Fig. 2.31).

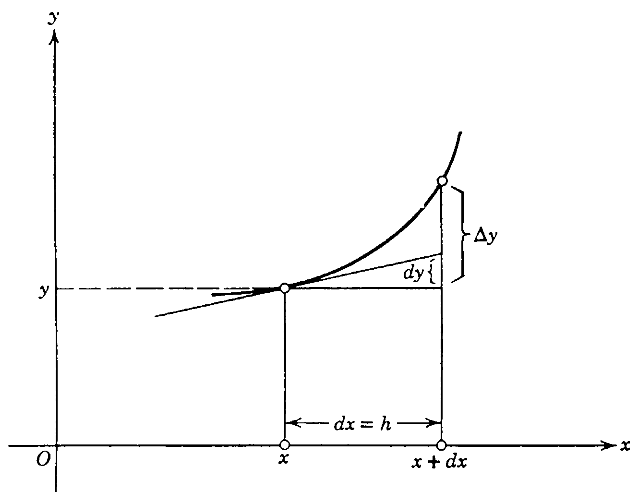


Figura 2.31 Incremento Δy y diferencial dy.

Aproximación lineal

Una estimación más precisa de la magnitud del *"error"*, esto es, de la desviación de la función $f(x)$ respecto de la función lineal que representa a la tangente, está dada mediante el teorema del valor medio del cálculo diferencial. Para una ξ adecuada entre x y $x+h$ se tiene

$$f(x+h) - f(x) = hf'(\xi),$$

de modo que

$$\varepsilon = \frac{f(x+h) - f(x)}{h} - f'(x) = f'(\xi) - f'(x).$$

Si, como ocurre usualmente en las aplicaciones, la función $f'(x)$ misma posee una derivada $f''(x)$, se encuentra, aplicando el teorema del valor medio por segunda vez, que

$$f'(\xi) - f'(x) = (\xi - x)f''(\eta),$$

donde η es un valor intermedio entre x y ξ y por lo tanto también entre x y $x + h$. Se sigue que

$$|\varepsilon| = |(\xi - x)f''(\eta)| = |\xi - x|\,|f''(\eta)| \leq hM,$$

donde M es una cota superior para el valor absoluto de la segunda derivada de f en el intervalo $[x, x + h]$. Entonces $|\varepsilon h|$, que mide la desviación de $f(x + h)$ respecto a la función lineal $f(x) + hf'(x)$, es a lo más igual a Mh^2. Para h suficientemente pequeño la expresión Mh^2 es, por supuesto, mucho menor que $f'(x)h$, a menos que suceda que $f'(x)$ posee el valor cero. Esta *aproximación de una función en un intervalo pequeño mediante una función lineal* es de gran significado tanto para las aplicaciones prácticas como para el análisis matemático avanzado. En capítulos posteriores volveremos a este tópico, y se deducirá entonces la mejor estimación de $|\varepsilon h| \leq \frac{1}{2}Mh^2$.

Interpolación

*Cuando una función $f(x)$ está descripta numéricamente mediante una tabla de valores, f está determinada ordinariamente mediante *interpolación lineal* para argumentos x intermedios entre aquéllos para los cuales f está registrada. Este procedimiento corresponde también a reemplazar la función f mediante una función lineal en un intervalo. En este caso la gráfica de la función lineal está dada mediante una *secante* en vez de una *tangente* a la curva que representa f. Si, digamos, los valores de f son conocidos en los puntos a y b, reemplazamos a $f(x)$ para un x intermedio por la expresión

$$\phi(x) = f(a) + (x - a)\,\frac{f(b) - f(a)}{b - a},$$

la cual es lineal en x y da los valores correctos de f en los puntos extremos $x = a$ y $x = b$ del intervalo (véase Fig. 2.32). Aquí, nuevamente mediante el uso del teorema del valor medio, puede estimarse el error en esta aproximación. Se tiene

$$\begin{aligned} f(x) - \phi(x) &= (x - a)\left[\frac{f(x) - f(a)}{x - a} - \frac{f(b) - f(a)}{b - a}\right] \\ &= (x - a)[f'(\xi_1) - f'(\xi_2)]. \end{aligned}$$

Puesto que ξ_1 está situado entre a y x, estará también situado entre a y b, como lo está ξ_2. Una segunda aplicación del teorema del valor medio del cálculo diferencial conduce entonces a

$$f(x) - \phi(x) = (x - a)f''(\eta)(\xi_1 - \xi_2),$$

donde η está entre ξ_1 y ξ_2 y, por lo tanto, también entre a y b. Consecuentemente, denotando por M a una cota superior para $|f''|$ en el intervalo $[a, b]$, se encuentra que

$$|f(x) - \phi(x)| \leq |x - a|\,|\xi_1 - \xi_2|\,|f''(\eta)| \leq M(b - a)^2.$$

Una vez más, la desviación de f de su aproximación lineal puede ser estimada mediante el cuadrado de la longitud del intervalo.

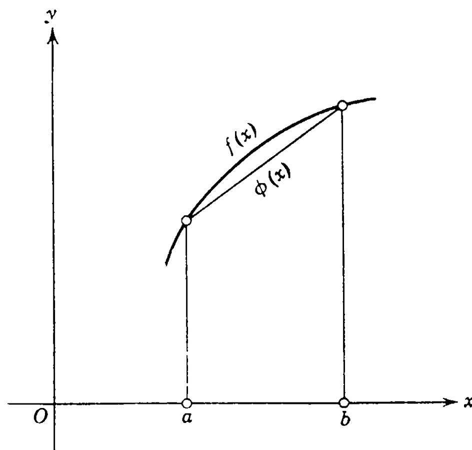


Figura 2.32 Interpolación lineal.

Como ejemplo numérico se toman de una tabla de funciones trigonométricas con medida en radianes los valores

$$\operatorname{sen} 0.75 = 0.6816, \qquad \operatorname{sen} 0.76 = 0.6889,$$

donde los errores no exceden de 0.00005. Si se desea deducir el valor de la función seno para el argumento intermedio 0.754, se encuentra mediante interpolación lineal que

$$\operatorname{sen} 0.754 \approx 0.6816 + \tfrac{4}{10}(0.6889 - 0.6816) \approx 0.6845.$$

Para la función $f(x) = \operatorname{sen} x$ la primera derivada es $f'(x) = \cos x$, y la segunda derivada $f''(x) = -\operatorname{sen} x$. Obviamente, $|f''(x)| \leq 1$, de modo que el error en el valor que se encontró para sen 0.754 como resultado del procedimiento de interpolación lineal no excede de $1 \times (0.01)^2 = 0.0001$. A esta estimación del error debemos añadir posiblemente errores debidos a redondeos en los valores tabulados y en la interpolación.

Puede compararse este valor obtenido mediante interpolación lineal con el valor que se obtendría reemplazando la curva seno por su tangente en el punto $x = 0.75$. Tomando $f'(0.75) = \cos 0.75 = 0.7317$ de la tabla, se encuentra

$$\operatorname{sen} 0.754 \approx f(0.75) + f'(0.75)(0.004) = \operatorname{sen} 0.75 + 0.004 \cos 0.75$$
$$\approx 0.6845.$$

Nótese que el valor verdadero de sen 0.754 correcto para seis cifras significativas es 0.684560.

k. Observaciones sobre aplicaciones a las ciencias naturales

Al aplicar las matemáticas a fenómenos naturales nunca se opera con cantidades *conocidas en forma precisa*. El que una longitud sea *exactamente* un metro es una cuestión que no puede ser decidida mediante ningún experimento y que en consecuencia no posee significado físico. Además, no existe significado físico al decir que la longitud de una varilla de material es racional o irracional; puede ser medida siempre con cualquier grado deseado de precisión mediante números racionales, y el único problema significativo es el de si se puede manipular para efectuar tal medición utilizando números racionales con denominadores relativamente pequeños. Tal como la cuestión de racionalidad o de irracionalidad en el sentido riguroso de "matemáticas exactas" no posee significado físico, efectuar procesos de límite en las aplicaciones no es usualmente más que una idealización matemática.

El significado práctico —y sobresaliente— de tales idealizaciones estriba en el hecho de que a través de idealizaciones las expresiones analíticas se tornan esencialmente más sencillas y más manejables. Por ejemplo, es mucho más simple y más conveniente operar con la noción de velocidad instantánea, la cual es función de *un* solo instante definido de tiempo, que con la noción de velocidad promedio entre dos instantes diferentes. Sin tal idealización toda investigación científica de la Naturaleza estaría condenada a complicaciones irremediables y se atollaría desde el principio.

No se tiene la intención de entrar en una discusión filosófica de la relación de las matemáticas con la realidad. Para un mejor entendimiento de la teoría se deberá destacar que en las aplicaciones se tiene el derecho de reemplazar una derivada por un cociente incremental y viceversa, siempre que las diferencias sean suficientemente pequeñas para garantizar una aproximación suficientemente cercana. El físico, el biólogo, el ingeniero, o cualquier otro que tenga que tratar con estas ideas en la práctica, tendrá por ello el derecho de identificar el cociente incremental con la derivada, dentro de sus límites de precisión. Mientras menor sea el incremento $h = dx$ de la variable independiente, con mayor precisión podrá representarse el incremento $\Delta y = f(x + h) - f(x)$ mediante la diferencial $dy = hf'(x)$. Mientras se mantenga, a sabiendas, dentro de los límites de precisión requeridos por el problema, puede estarle permitido hablar de las cantidades $dx = h$ y $dy = hf'(x)$ como "infinitesimales". Estas cantidades "físicamente

infinitesimales" poseen un significado preciso. Dichas cantidades son variables con valores finitos, diferentes de cero y elegidos suficientemente pequeños para la investigación dada; por ejemplo, menores que una parte fraccional de una longitud de onda, o menores que la distancia entre dos electrones en un átomo; en general, menores que el grado de precisión requerido.

2.9 La integral, la función primitiva y los teoremas fundamentales del cálculo

a. La derivada de la integral

Como se afirmó ya, la conexión entre integración y derivación es la piedra angular del cálculo diferencial e integral.

Recordamos de la sección 2.4 que una integral indefinida de una función continua $f(x)$ está definida como una función $\phi(x)$ del punto extremo superior de integración mediante la fórmula

$$\phi(x) = \int_{\alpha}^{x} f(u)\, du,$$

donde α era cualquier punto en el dominio de f. Se probará ahora:

El Teorema Fundamental del Cálculo (Primera Parte). *La integral indefinida $\phi(x)$ de una función continua $f(x)$ posee siempre una derivada $\phi'(x)$ y, además,*

$$\phi'(x) = f(x).$$

Esto es, la derivación de la integral indefinida de una función continua reproduce siempre el integrando:

$$\frac{d}{dx}\int_{\alpha}^{x} f(u)\, du = f(x).$$

Este carácter inverso de las operaciones de derivación e integración es el hecho básico del cálculo. La demostración es una consecuencia inmediata del teorema del valor medio del cálculo integral. De acuerdo con ese teorema, para cualesquiera valores de x y $x + h$ del dominio de f se tiene

$$\phi(x + h) - \phi(x) = \int_{x}^{x+h} f(u)\, du = hf(\xi),$$

donde ξ es algún valor en el intervalo con puntos extremos x y $x + h$. Para h tendiendo a cero el valor ξ debe tender a x, de modo que

$$\lim_{h\to 0} \frac{\phi(x + h) - \phi(x)}{h} = \lim_{h\to 0} f(\xi) = f(x),$$

puesto que f es continua. Por lo tanto, $\phi'(x) = f(x)$, como afirma el teorema.

Aplicaciones. (*a*) Puede utilizarse el teorema para encontrar derivadas de algunas funciones introducidas anteriormente. El logaritmo natural se definió para $x > 0$ mediante la integral indefinida

$$\log x = \int_1^x \frac{1}{u}\, du.$$

Se sigue inmediatamente que

$$\frac{d \log x}{dx} = \frac{1}{x}.$$

(*b*) Logaritmos más generales de una base arbitraria a fueron expresados en la forma

$$\log_a x = \frac{\log x}{\log a}.$$

Aplicando la regla para la derivada del producto de una constante y de una función se encuentra que

$$\frac{d}{dx} \log_a x = \frac{1}{x \log a}.$$

(*c*) Se encontró que

$$\frac{d}{dx} x^\alpha = \alpha x^{\alpha-1}$$

en el caso en que el exponente α es un entero o, más generalmente, un número racional. Puede ahora extenderse esta fórmula para α arbitrario. Para este propósito se recuerda la fórmula de integración

$$\int_a^b u^\beta\, du = \frac{1}{\beta + 1} (b^{\beta+1} - a^{\beta+1}),$$

la cual se probó para cualquier pareja de números positivos a, b y cualquier $\beta \neq -1$. Si se remplaza aquí el límite superior b por la variable x y se derivan ambos miembros respecto a x, se sigue, para $x > 0$, que

$$x^\beta = \frac{d}{dx} \frac{1}{\beta + 1} (x^{\beta+1} - a^{\beta+1}).$$

Utilizando las reglas para la derivada de una suma y el producto de una constante por una función, podemos escribir este resultado en la forma

$$x^\beta = \frac{1}{\beta + 1} \frac{d}{dx} x^{\beta+1}.$$

Sustituyendo α por $\beta + 1$, se obtiene la fórmula

$$\frac{d}{dx} x^\alpha = \alpha x^{\alpha-1}$$

para cualquier $\beta \neq -1$, esto es, para $\alpha \neq 0$. Sin embargo, la fórmula también se cumple trivialmente para $\alpha = 0$, pues entonces $x^\alpha = 1$, y la derivada de una constante es cero.

b. La función primitiva y su relación con la integral

Inversión de la derivación

El teorema fundamental muestra que la integral indefinida $\phi(x)$, esto es, la integral con un límite superior variable x, de una función $f(x)$, es una solución del siguiente problema: *Dada $f(x)$ determinar una función $F(x)$ tal que*

$$F'(x) = f(x).$$

Este problema requiere *invertir el proceso de derivación.* Es típico de los problemas inversos que ocurren en muchas partes de las matemáticas y que hemos encontrado ya como un método matemático fructífero para generar nuevos conceptos. (Por ejemplo, la primera extensión de la idea de los números naturales está sugerida por el deseo de invertir ciertos procesos elementales de la aritmética. Nuevamente se obtuvieron nuevos tipos de funciones de las inversas de funciones conocidas.)

Cualquier función $F(x)$ tal que $F'(x) = f(x)$, se denomina *función primitiva de $f(x)$*, o, simplemente, una *primitiva de $f(x)$*; esta terminología sugiere que la función $f(x)$ es deducida de $F(x)$.

Este problema de la inversión de la derivación, o de encontrar una función primitiva, es a primera vista de un carácter bastante diferente al problema de la integración. La primera parte del teorema fundamental del cálculo afirma, sin embargo:

Toda integral indefinida $\phi(x)$ de la función $f(x)$ es una primitiva de $f(x)$.

Empero, este resultado no resuelve completamente el problema de encontrar las funciones primitivas. Pues no se sabe aún si se han encontrado *todas* las soluciones del problema. La pregunta acerca del conjunto de todas las funciones primitivas es contestada por el siguiente teorema, mencionado en ocasiones como *segunda parte del teorema fundamental* del cálculo diferencial e integral

La diferencia de dos funciones primitivas $F_1(x)$ y $F_2(x)$ de la misma función $f(x)$ es siempre una constante

$$F_1(x) - F_2(x) = c.$$

Así, a partir de cualquier función primitiva $F(x)$ pueden obtenerse todas las demás en la forma

$$F(x) + c$$

mediante una elección adecuada de la constante c. Recíprocamente, para cada valor de la constante c la expresión $F_1(x) = F(x) + c$ *representa una función primitiva de* $f(x)$.

Es claro que para cualquier valor de la constante c la función $F(x) + c$ es una función primitiva, siempre que la propia $F(x)$ lo sea. Pues se tiene (ver p. 187)

$$\frac{d}{dx}[F(x) + c] = \frac{d}{dx}F(x) + \frac{d}{dx}c = F'(x) = f(x).$$

Por tanto, para completar la demostración del teorema queda sólo por mostrar que la diferencia de dos funciones primitivas $F_1(x)$ y $F_2(x)$ es siempre constante. Con este fin, considérese la diferencia

$$F_1(x) - F_2(x) = G(x).$$

Claramente,

$$G'(x) = F_1'(x) - F_2'(x) = f(x) - f(x) = 0.$$

Sin embargo, se probó en la p. 199, a partir del teorema del valor medio del cálculo diferencial, que una función cuya derivada se hace cero en todas partes de un intervalo es una constante. Por lo tanto, $G(x)$ es una constante c, y el teorema queda demostrado.

Combinando las dos partes que se acaban de probar, puede formularse ahora el

Teorema Fundamental del Cálculo. *Toda función primitiva* $F(x)$ *de una función dada* $f(x)$, *continua en un intervalo, puede ser representada en la forma*

$$F(x) = c + \phi(x) = c + \int_a^x f(u)\,du,$$

donde c y a son constantes; y, recíprocamente, para cualesquiera valores constantes de a y c, elegidos arbitrariamente,[1] *esta expresión representa siempre una función primitiva.*

Observaciones

Es posible conjeturar que la constante c puede, en general, ser omitida debido a que variando el límite inferior a se varía la función primitiva mediante una constante aditiva; esto es, que todas las funciones primitivas son integrales indefinidas. Frecuentemente, sin embargo, no pueden obtenerse todas las funciones primitivas si se omite la c, como el ejemplo $f(x) = 0$ muestra. Para esta función la integral indefinida será siempre cero, independientemente del límite inferior; así, cualquier constante arbi-

[1] Siempre que a esté situado en el dominio de f.

traria es una función primitiva de $f(x) = 0$. Un segundo ejemplo es la función $f(x) = \sqrt{x}$, la cual está definida sólo para valores no negativos de x. La integral indefinida es

$$\phi(x) = \tfrac{2}{3}x^{3/2} - \tfrac{2}{3}a^{3/2},$$

y se ve que no importa cómo se elija el límite inferior a; la integral indefinida $\phi(x)$ se obtiene siempre de $\frac{2}{3}(x)^{3/2}$ añadiendo una constante $-\frac{2}{3}a^{3/2}$ que es menor que o igual a cero; sin embargo, una función tal como $\frac{2}{3}x^{3/2} + 1$ es también una primitiva de $\sqrt{x}$. Así, en la expresión general para la función primitiva no puede hacerse caso omiso de la constante aditiva arbitraria.

La relación que se ha encontrado sugiere extender la noción de integral indefinida de modo que incluya todas las funciones primitivas. En lo sucesivo toda expresión de la forma $c + \phi(x) = c + \int_a^x f(u)\,du$ se denominará una *integral indefinida de* $f(x)$, *y no se distinguirá más entre la función primitiva y la integral indefinida.* Sin embargo, si el lector ha de poseer un correcto entendimiento de las interrelaciones de estos conceptos, es absolutamente necesario que tenga en mente que en primera instancia la integración y la inversión de la derivación son dos cosas diferentes, y que es sólo el conocimiento de la relación entre ellas lo que da el derecho de aplicar el término "integral indefinida" también a la función primitiva.

Es bastante usual utilizar una notación la cual no es perfectamente clara sin comentario: se escribe

$$F(x) = \int f(x)\,dx,$$

cuando se entiende que la función $F(x)$ es de la forma

$$F(x) = c + \int_a^x f(u)\,du$$

para constantes apropiadas c y a; esto es, se omite el límite superior x, el límite inferior a y la constante aditiva c, y se utiliza la letra x para la variable de integración. Estrictamente hablando, por supuesto, hay una pequeña incompatibilidad en utilizar la misma letra para la variable de integración y el límite superior de integración x que es la variable independiente en $F(x)$. Al usar la notación $\int f(x)\,dx$ no debe jamás perderse de vista la indeterminación asociada a ella, esto es, el hecho de que el símbolo denota siempre *una* de las funciones primitivas de f solamente. La fórmula $F(x) = \int f(x)\,dx$ es justamente una forma simbólica de escribir la relación

$$\frac{d}{dx}F(x) = f(x).$$

c. El uso de la función primitiva para la evaluación de integrales definidas

Supóngase que se conoce cualquier función primitiva $F(x)$ para la función $f(x)$ y que se desea evaluar la integral definida $\int_a^b f(u)\,du$. Se sabe que la integral indefinida

$$\phi(x) = \int_a^x f(u)\,du,$$

siendo también una primitiva de $f(x)$, puede sólo diferir de $F(x)$ por en constante aditiva. Por ello

$$\phi(x) = F(x) + c,$$

y la constante aditiva c está determinada de inmediato debido a que la integral indefinida $\phi(x) = \int_a^x f(u)\,du$ debe tomar el valor cero cuando $x = a$. Se obtiene así $0 = \phi(a) = F(a) + c$, de lo cual se sigue que $c = -F(a)$ y $\phi(x) = F(x) - F(a)$. En particular, para el valor $x = b$ se tiene la fórmula básica

$$\int_a^b f(u)\,du = F(b) - F(a),$$

si

$$F'(u) = f(u).$$

Por ello,

Si $F(x)$ es cualquier función primitiva de la función continua $f(x)$, la integral definida de $f(x)$ entre los límites a y b es igual a la diferencia $F(b) - F(a)$.

Si se utiliza la relación $F'(x) = f(x)$, esta consecuencia del teorema fundamental puede ser escrita en la forma

$$(31) \qquad F(b) - F(a) = \int_a^b F'(x)\,dx = \int_a^b \frac{dF(x)}{dx}\,dx = \int_a^b dF(x),$$

donde ahora $F(x)$ puede ser cualquier función con una derivada continua $F'(x)$, y donde se utiliza la sugestiva notación simbólica $dF(x) = F'(x)\,dx$ de Leibnitz.

Al aplicar nuestra regla, a menudo se utiliza una barra vertical para denotar la diferencia de valores en los puntos extremos, escribiéndose

$$\int_a^b \frac{dF(x)}{dx}\,dx = F(b) - F(a) = F(x)\Big|_a^b.$$

Puede escribirse (31) en la forma

$$\frac{F(b) - F(a)}{b - a} = \frac{1}{b - a}\int_a^b F'(x)\,dx. \tag{32}$$

Recordando la definición de la p. 163 relativa al promedio de una función en un intervalo, la regla afirma entonces que *el cociente incremental de la función $F(x)$ formado para los puntos a y b es igual a la media aritmética o promedio de la derivada de $F(x)$ en el intervalo con puntos extremos a y b.* Cuando se consideró el movimiento de una partícula sobre una recta, a la variación en distancia s dividida por la variación en tiempo t se le denominó "velocidad promedio". Se ve ahora que en realidad $\Delta s/\Delta t$ es precisamente el promedio de las velocidades ds/dt para el intervalo de tiempo dado, si t es la variable independiente utilizada al formar el promedio.

RELACIÓN ENTRE LOS TEOREMAS DE VALOR MEDIO

La fórmula

$$F(b) - F(a) = \int_a^b f(x)\,dx \tag{33}$$

que se cumple para cualquier función continua f y una de sus primitivas, F, hace también evidente la relación entre los teoremas del valor medio del cálculo integral (p. 163) y del cálculo diferencial (p. 194). Mediante el teorema del valor medio del cálculo integral se concluye de (33) que

$$F(b) - F(a) = (b - a)f(\xi).$$

Puesto que F es una primitiva de f, puede reemplazarse $f(\xi)$ por $F'(\xi)$ y obtenerse el teorema del valor medio del cálculo diferencial para la función F. Por supuesto, el requisito de que F posea derivada continua es más fuerte que el requisito del teorema del valor medio del cálculo diferencial, o sea, que la derivada simplemente exista.

d. Ejemplos

En el capítulo 3 se hará un uso extensivo del teorema fundamental al evaluar integrales. Por el momento se ilustra el método que está basado en la utilización de la fórmula

$$\int_a^b \frac{dF(x)}{dx} = F(b) - F(a)$$

mediante algunos ejemplos.

En la p. 185 se dedujo la fórmula

$$\frac{d}{dx}x^n = nx^{n-1}$$

para enteros positivos n. Esta fórmula es, en realidad, una consecuencia trivial del *teorema del binomio,* puesto que

$$\frac{d}{dx}x^n = \lim_{h\to 0}\frac{1}{h}[(x+h)^n - x^n]$$

$$= \lim_{h\to 0}\frac{1}{h}\left(x^n + nhx^{n-1} + \frac{n(n-1)}{2}h^2x^{n-2} + \cdots + h^n - x^n\right)$$

$$= \lim_{h\to 0}\left(nx^{n-1} + \frac{n(n-1)}{2}hx^{n-2} + \cdots + h^{n-1}\right) = nx^{n-1}.$$

Integrando entre los límites a y b se encuentra que

$$\int_a^b nx^{n-1}\,dx = b^n - a^n.$$

Escribiendo m para $n-1$ se obtiene la fórmula

$$\int_b^a x^m\,dx = \frac{1}{m+1}(b^{m+1} - a^{m+1})$$

para enteros $m \geq 0$. Esta deducción de la expresión para la integral de x^m es mucho más sencilla que la dada en la p. 153, la cual estuvo basada en una subdivisión geométrica del intervalo $[a, b]$; además, el resultado es ahora, en verdad, más general puesto que podemos excluir de la hipótesis el que a y b sean positivos.

Las fórmulas

$$\frac{d\,\text{sen}\,x}{dx} = \cos x, \qquad \frac{d\cos x}{dx} = -\text{sen}\,x$$

fueron obtenidas en la p. 186 aplicando los teoremas de adición para funciones trigonométricas y utilizando $\lim_{h\to 0}\left(\frac{\text{sen}\,h}{h}\right) = 1$. Integrando se obtiene inmediatamente

$$\int_a^b \cos x\,dx = \text{sen}\,b - \text{sen}\,a, \qquad \int_a^b \text{sen}\,x\,dx = \cos a - \cos b.$$

Nuevamente, esta deducción de las fórmulas de integración a partir del teorema fundamental es más sencilla que aquélla basada en la definición de la integral definida considerando ésta como límite de una suma.

Suplemento. La existencia de la integral definida de una función continua

Se tiene todavía que demostrar el hecho de que la integral de una función $f(x)$ entre los límites a y b $(a < b)$ existe siempre que $f(x)$ sea continua en el intervalo cerrado $[a, b]$. La demostración estará basada prin-

cipalmente en la continuidad uniforme de $f(x)$ (véase p. 65): para cualquier ε positivo dado los valores de f en dos puntos cualesquiera ξ y η del intervalo difieren en menos que ε si ξ y η están suficientemente próximos entre sí; y el grado de proximidad depende solamente de ε y es independiente de ξ, η; en otras palabras, existe un módulo uniforme de continuidad $\delta(\varepsilon)$ tal que $|f(\xi) - f(\eta)| < \varepsilon$ para cualesquiera valores ξ, η en $[a, b]$ para los cuales $|\xi - \eta| < \delta$.

La definición de integral como un límite de sumas requiere que se subdivida el intervalo $[a, b]$ en n partes mediante puntos sucesivos $x_0, x_1, \ldots,$ x_n, donde $x_0 = a$, $x_n = b$ y $x_0 < x_1 < \cdots < x_n$. Sea S_n un nombre para una particular subdivisión de $[a, b]$, de este tipo, en n celdas. La *fineza* de la subdivisión será medida mediante la longitud de la mayor de las celdas resultantes, esto es, el máximo de las cantidades $\Delta x_i = x_i - x_{i-1}$, la cual se denominará la "norma" de S_n. Debido a la continuidad uniforme de f, los valores de f en dos puntos cualesquiera de la misma celda difieren en menos que ε tan pronto como la norma de S_n sea menor que $\delta = \delta(\varepsilon)$. Una suma de aproximación basada en la subdivisión S_n se obtiene eligiendo un valor ξ_i en cada celda $[x_{i-1}, x_i]$ y formando

$$F_n = \sum_{i=1}^{n} f(\xi_i)\Delta x_i.$$

Debemos probar que para la sucesión de subdivisiones S_n con norma tendiendo a cero, las sumas F_n convergen hacia un límite, el que se denotará por $\int_a^b f(x)\,dx$, y que el valor de este límite no depende de la particular elección de las subdivisiones y de puntos intermedios ξ_i. Para llevar a cabo la demostración se comparan primero los valores F_n y F_N pertenecientes a dos subdivisiones S_n y S_N, donde la norma de S_n es menor que δ y donde la subdivisión S_N es un "refinamiento" de S_n; esto es, todos los puntos de subdivisión de S_n aparecen entre aquéllos de S_N. En una notación modificada apropiadamente se tiene

$$F_N = \sum_{j=1}^{N} f(\eta_j)\Delta y_j,$$

donde los valores y_j son los puntos de subdivisión de S_N, donde $\Delta y_j = y_j - y_{j-1}$, y η_j está situado en el intervalo $[y_{j-1}, y_j]$. Dos puntos de subdivisión sucesivos x_{i-1} y x_i de S_n aparecen también entre los valores y_j, por ejemplo, $x_{i-1} = y_{r-1}$, $x_i = y_s$. En S_N la celda $[x_{i-1}, x_i]$ es cortada en intervalos, digamos, $[y_{r-1}, y_r]$, $[y_r, y_{r+1}]$, $\ldots$, $[y_{s-1}, y_s]$, siendo

$$\sum_{j=r}^{s} f(\eta_j)(y_j - y_{j-1})$$

la contribución total a F_N. Se compara ésta con la contribución a F_n de la celda $[x_{i-1}, x_i]$, dada por $f(\xi_i)(x_i - x_{i-1})$, la cual puede ser escrita como

$$\sum_{j=r}^{s} f(\xi_i)(y_j - y_{j-1})$$

(véase Fig. 2.33), y se encuentra para el valor absoluto de la diferencia de las contribuciones:

$$\left|\sum_{j=r}^{s} [f(\eta_j) - f(\xi_i)](y_j - y_{j-1})\right| \leq \sum_{j=r}^{s} \varepsilon \cdot (y_j - y_{j-1}) = \varepsilon(x_i - x_{i-1}).$$

Por lo tanto, sumando las diferencias de las contribuciones a F_n y F_N para todas las celdas $[x_{i-1}, x_i]$ de S_n, se encuentra la estimación

$$|F_N - F_n| \leq \sum_{i=1}^{n} \varepsilon(x_i - x_{i-1}) = \varepsilon(b - a),$$

siempre que S_n posea una norma menor que $\delta(\varepsilon)$ y S_N sea un refinamiento de S_n.

Si ahora S_n y S_m son dos subdivisiones cualesquiera, puede considerarse la subdivisión S_N formada por todos los puntos de subdivisión de S_n junto con los de S_m. Entonces, S_N será un refinamiento tanto de S_n como de S_m.

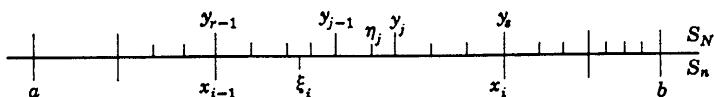


Figura 2.33.

Supóngase que S_n y S_m poseen norma menor que $\delta(\varepsilon)$. Eligiendo cualesquiera puntos intermedios η_j de las celdas de S_N para definir F_N, se encuentra

$$|F_n - F_m| = |(F_n - F_N) + (F_N - F_m)| \leq |F_n - F_N| + |F_m - F_N|$$
$$\leq 2\varepsilon(b - a).$$

Se ve entonces que dos sumas de aproximación cualesquiera difieren arbitrariamente poco una de la otra, si las normas de las correspondientes subdivisiones son suficientemente pequeñas. Considérese ahora cualquier sucesión de subdivisiones S_n cuyas normas tiendan a cero para $n \mapsto \infty$. Sean F_n las correspondientes sumas de aproximación. Para cualquier $\varepsilon > 0$ la norma de S_n es menor que $\delta(\varepsilon)$ para todo n suficientemente grande,. Por lo tanto

$$|F_n - F_m| < 2\varepsilon(b - a)$$

para n y m suficientemente grandes. Se sigue que la sucesión F_n satisface el criterio de convergencia de Cauchy (véase p. 120); consecuentemente,

$$\lim_{n\to\infty} F_n = F$$

existe.

Queda por probar que el valor de $\lim\limits_{n\to\infty} F_n$ no depende de las particulares subdivisiones y de los puntos intermedios. Si S_n' denota cualquier otra sucesión de subdivisiones cuyas normas tienden a cero, entonces la suma correspondiente F_n' posee un límite F'. Puesto que

$$|F_n' - F_n| < 2\varepsilon(b - a)$$

siempre que las normas de S_n y S_n' sean menores que $\delta(\varepsilon)$, se encuentra para $n \to \infty$ que también $|F - F'| \leq 2\varepsilon(b - a)$. Puesto que aquí ε es un número positivo arbitrario, se sigue que $F' = F$. Por lo tanto, el límite F, que se denota por $\int_a^b f(x)\,dx$, está determinado de manera única.

La demostración de la existencia de la integral definida de una función continua queda así completamente terminada.

Sumas de aproximación más generales. Nuestra demostración indica más claramente lo que es esencial en la aproximación de una integral mediante una suma. Se hace evidente el hecho de que se podría formular un proceso de límite algo más general que conduzca también a la integral, y que la siguiente forma, más general, del teorema es cierta: f_i no tiene que ser un valor de la función para que las sumas $F_n = \Sigma f_i\,\Delta x_i$ converjan a la integral; en vez de esto es suficiente que $|f_i - f(\xi_i)| < \delta(\varepsilon)$ para algún punto ξ_i en el intervalo $[x_{i-1}, x_i]$, donde $\delta(\varepsilon) \to 0$ para $\varepsilon \to 0$.

Este aserto general es útil a menudo. Si, por ejemplo, $f(x) = \phi(x)\psi(x)$, entonces en vez de la suma $\Sigma f(\xi_\nu)\,\Delta x_\nu$, puede considerarse la suma más general

$$\Sigma\, \phi(\xi_\nu')\,\psi(\xi_\nu'')\,\Delta x_\nu,$$

donde ξ_ν' y ξ_ν'' son dos puntos de la celda no necesariamente coincidentes. Esta suma también tiende a la integral

$$\int_a^b f(x)\,dx = \int_a^b \phi(x)\psi(x)\,dx$$

conforme n crece, siempre que la longitud de la máxima celda tienda a cero.

Una propiedad similar vale para otras sumas formadas de manera análoga; por ejemplo, la suma

$$\sum_{\nu=1}^{n} \sqrt{\phi(\xi_\nu')^2 + \psi(\xi_\nu'')^2}\,\Delta x_\nu$$

tiende a la integral

$$\int_a^b \sqrt{\phi(x)^2 + \psi(x)^2}\,dx.$$

Para probar estas propiedades se tiene solamente que probar que la variación D en las sumas de aproximación debida a la desviación de ξ_ν'' con respecto a ξ_ν' tiende a cero en el límite. Esto es obvio en el primer ejemplo, donde la variación en la suma de aproximación es

$$D = \sum_{\nu=1}^{n} \phi(\xi_\nu')[\psi(\xi_\nu'') - \psi(\xi_\nu')]\,\Delta x_\nu.$$

Puesto que ϕ es acotada y ψ uniformemente continua, D puede hacerse arbitrariamente pequeña eligiendo celdas suficientemente pequeñas.

La variación en la segunda suma está representada por

$$D = \sum_{\nu=1}^{n} \left(\sqrt{\phi(\xi_\nu')^2 + \psi(\xi_\nu'')^2} - \sqrt{\phi(\xi_\nu')^2 + \psi(\xi_\nu')^2}\right)\Delta x_\nu.$$

Utilizando la desigualdad del triángulo aplicada al triángulo con vértices $(a, 0)$, $(0, b)$, $(0, c)$ en la forma $|\sqrt{a^2 + b^2} - \sqrt{a^2 + c^2}| \leq |b - c|$, se encuentra que

$$|\sqrt{\phi(\xi_\nu')^2 + \psi(\xi_\nu'')^2} - \sqrt{\phi(\xi_\nu')^2 + \psi(\xi_\nu')^2}| \leq |\psi(\xi_\nu'') - \psi(\xi_\nu')|$$

de la que se sigue inmediatamente que D tiende a cero.

PROBLEMAS

SECCION 2.1, página 142

1. Sea f una función monótona positiva definida en $[a, b]$, donde $0 < a < b$. Sea ϕ la inversa de f y sean $\alpha = f(a)$, $\beta = f(b)$. Utilizando la interpretación de la integral como un área, mostrar que

$$\int_\alpha^\beta \phi(y)\,dy = b\beta - a\alpha - \int_a^b f(x)\,dx.$$

SECCION 2.2, página 150

1. Probar para cualquier número natural p que

$$\int_a^b x^p\,dx = \frac{1}{p+1}(b^{p+1} - a^{p+1}),$$

utilizando una subdivisión de $[a, b]$ en celdas de igual longitud. Emplear las técnicas del capítulo 1, problemas diversos 5 a 12, para evaluar las sumas de aproximación F_n.

2. Deducir la fórmula para $\int_a^b x^\alpha\,dx, a, b > 0$, donde α es racional y negativo; por ejemplo, $\alpha = -r/s$, siendo r y s números naturales. *(Sugerencia:* Hágase $q^{-1/s} = \tau$, donde $q = \sqrt[n]{b/a}$.)

3. Por el método utilizado para encontrar la integral de sen x, deducir la fórmula

$$\int_a^b \cos x\,dx = \operatorname{sen} b - \operatorname{sen} a.$$

4. Hacer una aseveración general relativa a $\int_{-a}^a f(x)\,dx$ cuando $f(x)$ es: (a) una función impar, y (b) una función par.

5. Calcular $\int_0^{\pi/2} \operatorname{sen} x\, dx$ y $\int_0^{\pi/2} \cos x\, dx$. Explicar con argumentos geométricos por qué éstas deben ser iguales. Además, explicar por qué

$$\int_a^{a+2\pi} \operatorname{sen} x\, dx = \int_b^{b+2\pi} \cos x\, dx$$

para todos los valores de a y b.

6. (*a*) Evaluar $I_n = \int_0^a x^{1/n}\, dx$. ¿Qué es $\lim\limits_{n\to\infty} I_n$? Interpretar geométricamente.

(*b*) Hacer lo mismo para $I_n = \int_0^a x^n\, dx$.

7. Evaluar

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}}\left(1 + \frac{1}{\sqrt{2}} + \cdots + \frac{1}{\sqrt{n}}\right).$$

SECCION 2.3, página 158

***1.** *Desigualdad de Cauchy para integrales.* Probar que para todas las funciones continuas $f(x)$, $g(x)$, es

$$\int_a^b [f(x)]^2\, dx \int_a^b [g(x)]^2\, dx \geq \left(\int_a^b f(x)g(x)\, dx\right)^2.$$

***2.** Probar que si $f(x)$ es continua y

$$f(x) = \int_0^x f(t)\, dt,$$

entonces $f(x)$ es idénticamente cero.

***3.** Sea $f(x)$ continua en el sentido de Lipschitz en [0, 1]; esto es

$$|f(x) - f(y)| < M\,|x - y|$$

para todo x, y en el intervalo. Probar que

$$\left|\int_0^1 f(x)\, dx - \frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k}{n}\right)\right| < \frac{M}{2n}$$

SECCION 2.5, página 167

1. Probar que

$$\log\frac{p}{q} \leq \frac{p-q}{\sqrt{pq}} \qquad (q \leq p).$$

(Sugerencia: Aplicar la desigualdad de *Cauchy,* Problema 1.)

2. (*a*) Verificar que $\log(1 + x) = \int_0^x \frac{1}{1+u}\, du$, donde $x > -1$.
(*b*) Mostrar para $x > 0$ que

$$x - \frac{x^2}{2} < \log(1 + x) < x.$$

*(*c*) Más generalmente, mostrar para $0 < x < 1$ que

$$x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots - \frac{x^{2n}}{2n} < \log(1 + x) < x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + \frac{x^{2n+1}}{2n+1}.$$

(*Sugerencia:* Comparar $1/(1+u)$ con una progresión geométrica.)

SECCION 2.6, página 171

1. (*a*) Probar que

$$\int_a^b e^x\,dx = e^b - e^a$$

utilizando una subdivisión de $[a, b]$ en celdas iguales. *Sugerencia:* Aplicar también $\log \alpha = \lim_{n\to\infty} n(\sqrt[n]{\alpha} - 1)$.

(*b*) Encontrar $\int_a^b \log x\,dx$. (Véase sección 2.1, problema 1.)

(*c*) Mostrar para $x \geq 0$ que

$$1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} \leq e^x \leq 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \frac{e^x x^{n+1}}{(n+1)!}.$$

(*Sugerencia:* Obtener estimaciones superiores e inferiores para $\int_0^x e^u\,du$ e integrar repetidamente.)

Obtener estimaciones del mismo tipo para e^x cuando $x < 0$.

SECCION 2.8c, página 184

Calcular las derivadas de las siguientes funciones dondequiera que estén definidas, directamente como el límite de sus cocientes de incrementos.

1. $\tan x$.
2. $\sec^2 x$.
3. $\operatorname{sen} \sqrt{x}$.
4. $\sqrt{\operatorname{sen} x}$.
5. $\dfrac{1}{\operatorname{sen} x}$
6. $\operatorname{sen} \dfrac{1}{x}$
7. x^α, donde α es racional y negativo.

SECCION 2.8i, página 196

1. Mostrar que $x > \operatorname{sen} x$ para x positivo y $x < \tan x$ para x en $\left(0, \dfrac{\pi}{2}\right)$.

2. Si $f(x)$ es continua y derivable para $a \leq x \leq b$, mostrar que si $f'(x) \leq 0$ para $a \leq x < \xi$ y $f'(x) \geq 0$ para $\xi < x \leq b$, la función nunca es menor que $f(\xi)$.

*3. Si la función continua $f(x)$ posee derivada $f'(x)$ en cada punto x de una vecindad de $x = \xi$, y si $f'(x)$ se aproxima a un límite L conforme $x \to \xi$, entonces $f'(\xi)$ existe y es igual a L.

*4. Sea $f(x)$ definida y derivable en todo el eje x. Mostrar que si $f(0) = 0$ y en todas partes es $|f'(x)| \leq |f(x)|$, entonces $f(x) = 0$ idénticamente.

SECCION 2.9, página 206

*1. Si una partícula recorre una distancia 1 en un tiempo 1, partiendo del reposo y terminando también en reposo, entonces en algún punto del intervalo deberá haber estado sujeta a una aceleración igual a 4 o más.

SUPLEMENTO, página 213. Existencia de la Integral Indefinida.

1. Sea $f(x)$ definida y acotada en $[a, b]$. Se definen la suma superior Σ y la suma inferior σ para la subdivisión

$$a = x_0 < x_1 < x_2 \cdots < x_n = b$$

como

$$\Sigma = \sum_{i=1}^{n} M_i \, \Delta x_i, \qquad \sigma = \sum_{i=1}^{n} m_i \, \Delta x_i,$$

donde M_i es el extremo superior y m_i es el extremo inferior de $f(x)$ en la celda $[x_{i-1}, x_i]$.

(*a*) Mostrar que en cualquier refinamiento de una subdivisión la suma superior decrece o bien permanece invariable; y que, análogamente, la suma inferior crece o bien permanece invariable.

(*b*) Probar que cada suma superior es mayor que o igual a cada suma inferior.

(*c*) La *integral superior de Darboux* F^+ está definida como el extremo inferior de las sumas superiores, y la *integral inferior de Darboux* F^-, como el extremo superior de las sumas inferiores, sobre todas las subdivisiones. De (*b*), $F^+ \geq F^-$. Si $F^+ = F^-$, al valor común se le denomina *integral de Darboux* de f. Probar que la integral de Darboux de f es en realidad la integral de Riemann ordinaria; además, mostrar que la integral de Riemann existe si y sólo si las integrales superior e inferior de Darboux existen y son iguales.

2. Sea $f(x)$ una función monótona definida en $[a, b]$.

(*a*) Mostrar que la diferencia entre las sumas superior e inferior para una subdivisión en n celdas iguales está dada exactamente por

$$\Sigma - \sigma = |f(b) - f(a)| \, (b - a)/n.$$

y explicar este resultado geométricamente.

(*b*) Usar el resultado de (*a*) para probar que la integral de Darboux existe.

(*c*) Estimar $\Sigma - \sigma$ en términos de $f(a)$, $f(b)$ y la amplitud de la subdivisión si las celdas de la subdivisión pueden ser desiguales.

(*d*) En la mayor parte de los casos $f(x)$, si no es monótona, puede escribirse como la suma de funciones monótonas, $f(x) = \phi(x) + \psi(x)$, donde ϕ es no creciente y ψ no decreciente. Estimar las diferencias entre las sumas superior e inferior en ese caso.

3. Mostrar que si $f(x)$ posee derivada continua en el intervalo cerrado $[a, b]$, entonces $f(x)$ puede escribirse como la suma de funciones monótonas como en el problema 2*d*.

PROBLEMAS DIVERSOS

1. Probar que

$$(a) \int_{-1}^{1} (x^2 - 1)^2 \, dx = \frac{16}{15}; \qquad (b) \ (-1)^n \int_{-1}^{1} (x^2 - 1)^n \, dx = \frac{2^{2n+1}(n!)^2}{(2n+1)!}$$

2. Probar para el coeficiente binomial $\binom{n}{k}$ que

$$\binom{n}{k} = \left[(n+1) \int_0^1 x^k (1-x)^{n-k}\, dx \right]^{-1}.$$

*3. Si $f(x)$ posee una derivada $f'(x)$ (no necesariamente continua) en cada punto x de $a \leq x \leq b$, y si $f'(x)$ toma los valores m y M, también toma cualquier valor μ entre m y M.

4. Si $f''(x) \geq 0$ para todos los valores de x en $a \leq x \leq b$, la gráfica de $y = f(x)$ está situada sobre o por encima de la línea tangente en cualquier punto $x = \xi$, $y = f(\xi)$ de la gráfica.

5. Si $f''(x) \geq 0$ para todos los valores de x en $u \leq x \leq b$, la gráfica de $y = f(x)$ en el intervalo $x_1 \leq x \leq x_2$ está situada por debajo del segmento de recta que une los dos puntos de la gráfica para los cuales $x = x_1$, $x = x_2$.

6. Si $f''(x) \geq 0$, entonces $f\left(\dfrac{x_1 + x_2}{2}\right) \leq \dfrac{f(x_1) + f(x_2)}{2}$.

*7. Sea $f(x)$ una función tal que $f''(x) \geq 0$ para todos los valores de x y sea $u = u(t)$ una función continua arbitraria. Entonces

$$\frac{1}{a}\int_0^a f[u(t)]\, dt \geq f\left(\frac{1}{a}\int_0^a u(t)\, dt\right).$$

8. (*a*) Derivar directamente y escribir abajo las fórmulas correspondientes de integración: (i) $x^{1/2}$; (ii) $\tan x$.
(*b*) Evaluar

$$\lim_{n\to\infty} \frac{1}{n}\left(1 + \sec^2\frac{\pi}{4n} + \sec^2\frac{2\pi}{4n} + \cdots + \sec^2\frac{n\pi}{4n}\right).$$

9. Sea $f(x)$ una función que posee derivadas primera y segunda para todos los valores reales de x. Probar que si $f(x)$ es positiva en todas partes y cóncava, entonces $f(x)$ es constante.

3

Las técnicas del cálculo

Parte A Derivación e integración de las funciones elementales

3.1 Las reglas más simples para derivar, y sus aplicaciones

Aún cuando los problemas de integración son usualmente de mayor importancia que los de diferenciación, estos últimos ofrecen menor dificultad formal que los primeros. Por ello es natural el procedimiento de conocer a fondo el arte de derivar la más amplia clase posible de funciones; entonces, por el teorema fundamental (sección 2.9), los resultados de la derivación estarán disponibles para la evaluación de integrales. En las siguientes secciones continuaremos tales aplicaciones del teorema fundamental. Hasta cierto punto se hará un nuevo comienzo y se desarrollarán las técnicas de integración sistemáticamente sobre la base de ciertas reglas generales para la diferenciación.

a. Reglas de la derivación

Se supone que en el intervalo bajo consideración las funciones $f(x)$ y $g(x)$ son derivables; entonces las siguientes reglas son básicas.

Regla 1. Multiplicación por una constante. Para cualquier constante c, la función $\phi(x) = cf(x)$ es derivable, y es

$$\phi'(x) = cf'(x) \tag{1}$$

La demostración, trivial, fue dada en el capítulo 2, p. 187.

Regla 2. Derivada de una suma. Si $\phi(x) = f(x) + g(x)$, entonces $\phi(x)$ es derivable y es

$$\phi'(x) = f'(x) + g'(x); \tag{2}$$

esto es, las operaciones de derivación y adición son intercambiables. Lo mismo se cumple para la suma de un número finito n de funciones derivables,

$$\phi(x) = \sum_{\nu=1}^{n} f_\nu(x),$$

para la cual se obtiene

$$\phi'(x) = \sum_{\nu=1}^{n} f_\nu'(x).$$

La demostración es obvia a partir de la definición de derivada.

Regla 3. Derivada de un producto. Si $\phi(x) = f(x)g(x)$, entonces $\phi(x)$ es derivable y es

$$\phi'(x) = f(x)g'(x) + g(x)f'(x). \tag{3}$$

La demostración se sigue a partir de la ecuación

$$\frac{\phi(x+h) - \phi(x)}{h} = \frac{f(x+h)g(x+h) - f(x)g(x)}{h}$$

$$= f(x+h)\,\frac{g(x+h) - g(x)}{h} + g(x)\,\frac{f(x+h) - f(x)}{h}.$$

Tomando el límite en esta expresión para $h \to 0$ se obtiene la Ec. (3).

Esta fórmula se vuelve más clara si se divide[1] por $\phi(x) = f(x)g(x)$. Se obtiene entonces

$$\frac{\phi'(x)}{\phi(x)} = \frac{f'(x)}{f(x)} + \frac{g'(x)}{g(x)}.$$

Utilizando la notación de diferenciales (capítulo 2, p. 200), puede también reescribirse la Ec. (3) como

$$d(fg) = f\,dg + g\,df.$$

Por inducción se obtiene para la derivada de un producto de n factores una expresión que consiste de n términos, cada uno de los cuales consta de la derivada de un factor multiplicada por todos los otros factores del producto original:

$$\begin{aligned}\phi'(x) &= \frac{d}{dx}[f_1(x)f_2(x)\cdots f_n(x)]\\ &= f_1'(x)f_2(x)\cdots f_n(x) + f_1(x)f_2'(x)f_3(x)\cdots f_n(x)\\ &\quad + \cdots + f_1(x)f_2(x)\cdots f_n'(x)\\ &= \sum_{\nu=1}^{n} f_\nu'(x)\,\frac{\phi(x)}{f_\nu(x)},\end{aligned}$$

---

[1] Debe, por supuesto, suponerse que $\phi(x)$ en ninguna parte es igual a cero.

o bien, dividiendo por $\phi(x) = f_1(x)f_2(x)\cdots f_n(x)$,

$$\frac{\phi'(x)}{\phi(x)} = \frac{f_1'(x)}{f_1(x)} + \frac{f_2'(x)}{f_2(x)} + \cdots + \frac{f_n'(x)}{f_n(x)} = \sum_{\nu=1}^{n} \frac{f_\nu'(x)}{f_\nu(x)},$$

la cual es válida donde se tenga $\phi(x) \neq 0$.

Mediante la aplicación repetida de la regla para la derivada de un producto pueden obtenerse también *fórmulas para la segunda derivada y derivadas de órdenes superiores.* Para la segunda derivada se tiene

$$\begin{aligned}\frac{d^2fg}{dx^2} &= \frac{d}{dx}\left(\frac{dfg}{dx}\right) = \frac{d}{dx}\left(f\frac{dg}{dx} + \frac{df}{dx}g\right)\\ &= \frac{d}{dx}\left(f\frac{dg}{dx}\right) + \frac{d}{dx}\left(\frac{df}{dx}g\right)\\ &= f\frac{d^2g}{dx^2} + 2\frac{df}{dx}\frac{dg}{dx} + \frac{d^2f}{dx^2}g.\end{aligned}$$

Regla de Leibnitz. El lector deberá probar por inducción que la n-ésima derivada de un producto puede encontrarse de acuerdo con la siguiente regla (regla de Leibnitz):

$$\begin{aligned}\frac{d^n}{dx^n}(fg) = f\frac{d^ng}{dx^n} &+ \binom{n}{1}\frac{df}{dx}\frac{d^{n-1}g}{dx^{n-1}}\\ &+ \binom{n}{2}\frac{d^2f}{dx^2}\frac{d^{n-2}g}{dx^{n-2}} + \cdots + \binom{n}{n-1}\frac{d^{n-1}f}{dx^{n-1}}\frac{dg}{dx} + \frac{d^nf}{dx^n}g.\end{aligned}$$

Aquí $\binom{n}{1} = n$, $\binom{n}{2} = [n(n-1)]/2!$, etc., denotan los coeficientes binomiales.

Regla 4. Derivada de un cociente. Para un cociente

$$\phi(x) = \frac{f(x)}{g(x)}$$

se cumple la siguiente regla: La función $\phi(x)$ es derivable en cada punto en el cual $g(x)$ no se hace cero, y

$$\phi'(x) = \frac{g(x)f'(x) - g'(x)f(x)}{[g(x)]^2} \tag{4}$$

Si $\phi(x) \neq 0$, esto puede ser escrito como

$$\frac{\phi'(x)}{\phi(x)} = \frac{f'(x)}{f(x)} - \frac{g'(x)}{g(x)}$$

DEMOSTRACIÓN. Si se supone la derivabilidad de $\phi(x)$, puede aplicarse la regla de un producto a $f(x) = \phi(x)g(x)$ y concluir que

$$f'(x) = \phi(x)g'(x) + g(x)\phi'(x).$$

Mediante la substitución de $f(x)/g(x)$ en vez de $\phi(x)$ en la derecha y despejando $\phi'(x)$, se obtiene la regla 4.

Puede probarse la derivabilidad de $\phi(x)$ así como la regla, si se escribe

$$\frac{\phi(x+h) - \phi(x)}{h} = \frac{\dfrac{f(x+h)}{g(x+h)} - \dfrac{f(x)}{g(x)}}{h}$$

$$= \frac{g(x)\dfrac{f(x+h) - f(x)}{h} - \dfrac{g(x+h) - g(x)}{h}f(x)}{g(x)g(x+h)}.$$

Si ahora se hace h tender a cero se obtiene el resultado establecido, pues, por hipótesis, el denominador no tiende a cero sino al límite $[g(x)]^2$, y los dos términos del numerador poseen límites $g(x)f'(x)$ y $g'(x)f(x)$ respectivamente. Esto prueba tanto la existencia del límite en el miembro izquierdo como la fórmula de derivación.

b. Derivación de las funciones racionales

Primero, se deduce una vez más la fórmula

$$\frac{d}{dx}x^n = nx^{n-1}$$

para todo entero positivo n, invocando la regla para derivar un producto. Considérese x^n como un producto de n factores, $x^n = x \cdots x$, y obténgase así

$$\frac{d}{dx}x^n = 1 \cdot x^{n-1} + 1 \cdot x^{n-1} + \cdots + 1 \cdot x^{n-1} = nx^{n-1}.$$

La segunda derivada de la función x^n se sigue a partir de esta fórmula y la Ec. (1):

$$\frac{d^2}{dx^2}x^n = n(n-1)x^{n-2}.$$

Continuando, se obtienen las derivadas de órdenes superiores

$$\frac{d^3}{dx^3}x^n = n(n-1)(n-2)x^{n-3}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\frac{d^n}{dx^n}x^n = 1 \cdot 2 \cdots n = n!.$$

A partir de la última de estas fórmulas es claro que la n-ésima derivada de x^n es una constante, mientras que la $(n + 1)$-ésima derivada se hace cero en todas partes.

Utilizando las dos primeras reglas y la regla para derivar potencias puede derivarse cualquier polinomio $y = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$, obteniéndose

$$y' = a_1 + 2a_2x + 3a_3x^2 + \cdots + na_nx^{n-1};$$

además,

$$y'' = 2a_2 + 3 \cdot 2a_3x + 4 \cdot 3a_4x^2 + \cdots + n(n-1)a_nx^{n-2},$$

y así sucesivamente.

La derivada de cualquier función racional puede encontrarse ahora con la ayuda de la regla del cociente. En particular, se deduce nuevamente la fórmula de derivación para la función x^n, donde $n = -m$ es un entero negativo. La aplicación de la regla del cociente, junto con el hecho de que la derivada de una constante es igual a cero, da el resultado

$$(d/dx)(1/x^m) = -mx^{m-1}/x^{2m} = -m/x^{m+1};$$

o bien, si se toma $m = -n$,

$$\frac{d}{dx}x^n = nx^{n-1},$$

que coincide formalmente con el resultado para valores positivos de n y con los resultados dados anteriormente (p. 186).

c. Derivación de las funciones trigonométricas

Para las funciones trigonométricas $\operatorname{sen} x$ y $\cos x$ se han obtenida ya (p. 186) las fórmulas de derivación

$$\frac{d}{dx}\operatorname{sen} x = \cos x \qquad \text{y} \qquad \frac{d}{dx}\cos x = -\operatorname{sen} x.$$

La regla del cociente nos permite ahora derivar las funciones

$$y = \tan x = \frac{\operatorname{sen} x}{\cos x} \qquad \text{y} \qquad y = \cot x = \frac{\cos x}{\operatorname{sen} x}.$$

De acuerdo con la regla, la derivada de la primera de estas funciones es

$$y' = \frac{\cos^2 x + \operatorname{sen}^2 x}{\cos^2 x} = \frac{1}{\cos^2 x},$$

de modo que

$$\frac{d}{dx}\tan x = \frac{1}{\cos^2 x} = \sec^2 x = 1 + \tan^2 x.$$

Análogamente, se obtiene

$$\frac{d}{dx}\cot x = -\frac{1}{\operatorname{sen}^2 x} = -\operatorname{cosec}^2 x = -(1 + \cot^2 x).$$

A las fórmulas de derivación para $\operatorname{sen} x$, $\cos x$, $\tan x$ y $\cot x$ corresponden las siguientes fórmulas de integración:

$$\int \cos x \, dx = \operatorname{sen} x, \qquad \int \operatorname{sen} x \, dx = -\cos x,$$

$$\int \frac{1}{\cos^2 x} dx = \tan x, \qquad \int \frac{1}{\operatorname{sen}^2 x} dx = -\cot x.$$

A partir de estas fórmulas se obtiene, por medio de la regla fundamental de la sección 2.9, p. 211, el valor de la integral definida, entre límites cualesquiera, siendo la única restricción el hecho de que cuando las dos últimas fórmulas son utilizadas el intervalo de integración no debe contener punto alguno de discontinuidad del integrando, tal como un múltiplo impar de $\pi/2$ en el primer caso y un múltiplo par de $\pi/2$ en segundo. Por ejemplo,

$$\int_a^b \cos x \, dx = \operatorname{sen} x \Big|_a^b = \operatorname{sen} b - \operatorname{sen} a.$$

3.2 La derivada de la función inversa

a. Fórmula general

Hemos visto en la p. 69 que una función continua $y = f(x)$ posee una inversa continua en todo intervalo en el cual aquélla es monótona. Precisamente:

Si $a \leq x \leq b$ es un intervalo en el cual la función continua $y = f(x)$ es monótona, y si $f(a) = \alpha$ y $f(b) = \beta$, entonces f posee una función inversa que es continua y monótona en el intervalo entre α y β.

Como se hizo notar en la p. 198, el signo de la derivada proporciona un criterio simple para ver cuándo una función es monótona y posee por consiguiente una inversa. Una función derivable es continua y será monótona creciente en un intervalo donde $f'(x)$ sea mayor que cero, y monótona decreciente en un intervalo en el cual $f'(x)$ sea en todas partes menor que cero.

Se caracterizará ahora la derivada de la función inversa probando el siguiente teorema.

TEOREMA. *Si la función $y = f(x)$ es derivable en el intervalo $a < x < b$ y si $f'(x) > 0$, o bien $f'(x) < 0$, en todo el intervalo, entonces la función*

inversa $x = \phi(y)$ posee también una derivada en cada punto interior de su intervalo de definición: la derivada de $y = f(x)$ y la de su inversa $x = \phi(y)$ satisfacen la relación $f'(x) \cdot \phi'(y) = 1$ en valores x, y correspondientes.

Esta relación puede también expresarse en la forma

$$\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}}. \tag{5}$$

Esta última fórmula nuevamente ilustra lo adecuado de la notación de Leibnitz: el cociente simbólico dy/dx puede ser tratado en las fórmulas como si fuera realmente una fracción.

DEMOSTRACIÓN. La demostración del teorema es sencilla. Escribiendo la derivada como el límite de un cociente de incrementos, se tiene

$$y' = f'(x) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{x_1 \to x} \frac{y_1 - y}{x_1 - x},$$

donde x e $y = f(x)$, y x_1 e $y_1 = f(x_1)$, respectivamente denotan parejas de valores correspondientes. Por hipótesis el primero de estos valores límite no es igual a cero. Debido a la continuidad de $y = f(x)$ y $x = \phi(y)$, las relaciones $y_1 \to y$ y $x_1 \to x$ son equivalentes. Por tanto, el valor límite

$$\lim_{x_1 \to x} \frac{x_1 - x}{y_1 - y} = \lim_{y_1 \to y} \frac{x_1 - x}{y_1 - y}$$

existe y es igual a $1/f'(x)$. Por otra parte, el valor límite en el miembro derecho es por definición la derivada $\phi'(y)$ de la función inversa $\phi(y)$, y nuestra fórmula queda probada.

El sencillo significado geométrico de la fórmula está mostrado claramente en la Fig. 3.1. La tangente a la curva $y = f(x)$, o bien $x = \phi(y)$, forma un ángulo α con el eje positivo x y un ángulo β con el eje positivo y; y, a partir de la interpretación geométrica de la derivada de una función como la pendiente de la tangente, se tiene

$$f'(x) \tan \alpha, \qquad \phi'(y) = \tan \beta.$$

Puesto que la suma de los ángulos α y β es $\pi/2$, $\tan \alpha \tan \beta = 1$, y esta relación es exactamente equivalente a nuestra fórmula de derivación.

Puntos críticos

Hasta ahora se ha supuesto expresamente que $f'(x) > 0$, o bien $f'(x) < 0$, esto es, que $f'(x)$ nunca es cero. ¿Qué sucede entonces si $f'(x) = 0$? Si $f'(x) = 0$ en todas partes de un intervalo, entonces f es

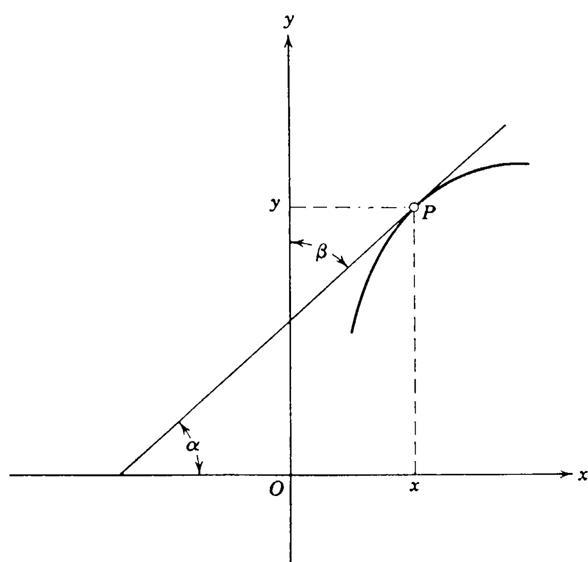


Figura 3.1 Derivación de la función inversa.

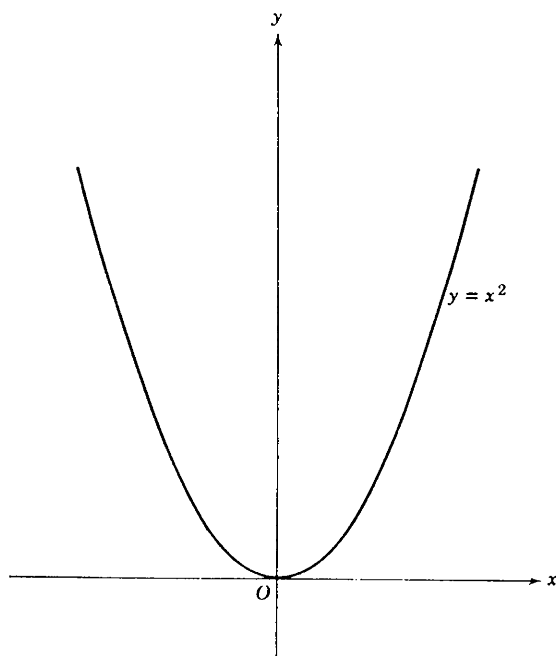


Figura 3.2 Parábola.

constante en él, y consecuentemente no posee inversa debido a que el mismo valor de y corresponde a todos los valores de x en el intervalo. Si $f'(x) = 0$ solamente en puntos "críticos" aislados (y si $f'(x)$ se supone continua), entonces se tienen dos casos, según que al pasar por estos puntos $f'(x)$ cambie o no de signo. En el primer caso este punto separa una región donde la función es monótona creciente de otra donde es monótona decreciente. En la vecindad de uno de tales puntos no puede haber función inversa univaluada. En el segundo caso, el que la derivada se haga cero no con-

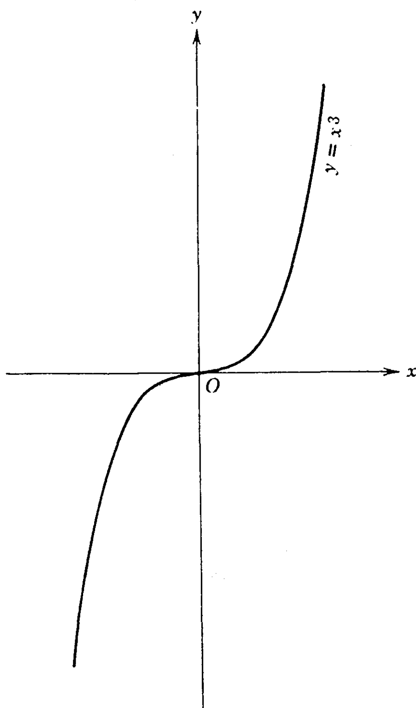


Figura 3.3 Parábola cúbica.

tradice el carácter monótono de la función $y = f(x)$, de modo que existe una función inversa univaluada. Sin embargo, la función inversa ya no sigue siendo derivable en el correspondiente punto; de hecho, la derivada es allí infinita. Las funciones $y = x^2$ e $y = x^3$ en el punto $x = 0$ ofrecen ejemplos de los dos tipos. La Fig. 3.2 y la Fig. 3.3 ilustran el comportamiento de las dos funciones al pasar por el origen, y al mismo tiempo muestran que la función $y = x^3$ posee una inversa univaluada, mientras que la otra función, $y = x^2$, no.

b. La inversa de la n-ésima potencia: la n-ésima raíz

El ejemplo más sencillo lo constituye la inversa de la función $y = x^n$ para enteros positivos n; primero se suponen valores positivos de x, por lo tanto también $y > 0$. Bajo estas condiciones y' es siempre positivo, de modo que para todos los valores positivos de y puede formarse la función inversa única

$$x = \sqrt[n]{y} = y^{1/n}.$$

La derivada de esta función inversa se obtiene inmediatamente por la regla general anterior como sigue:

$$\frac{d(y^{1/n})}{dy} = \frac{dx}{dy} = \frac{1}{dy/dx} = \frac{1}{nx^{n-1}} = \frac{1}{n}\frac{1}{y^{(n-1)/n}} = \frac{1}{n}y^{(1/n)-1}.$$

Si se cambia ahora la notación y se denota nuevamente la variable independiente por x, puede finalmente escribirse

$$\frac{d\sqrt[n]{x}}{dx} = \frac{d}{dx}(x^{1/n}) = \frac{1}{n}x^{(1/n)-1},$$

que coincide con el resultado obtenido en la p. 185.

Para $n > 1$, el punto $x = 0$ requiere una consideración especial. Si x se aproxima a cero a través de valores positivos, $d(x^{1/n})/dx$ crecerá obviamente más allá de toda cota; esto corresponde al hecho de que para $n > 1$ la derivada de la n-ésima potencia $f(x) = x^n$ se hace cero en el origen. Geométricamente esto significa que las curvas $y = x^{1/n}$ para $n > 1$ tocan el eje y en el origen (véase Fig. 1.35, p. 72).

Se deberá notar que para valores impares de n la suposición $x > 0$ puede omitirse y la función $y = x^n$ es monótona y posee una inversa sobre todo el dominio de números reales. La fórmula

$$\frac{d(\sqrt[n]{y})}{dy} = (1/n)y^{(1/n)-1}$$

se sigue cumpliendo para valores negativos de y, pero para $x = 0$, con $n > 1$, se tiene $d(x^n)/dx = 0$, que corresponde a una derivada infinita dx/dy de la función inversa en el punto $y = 0$.

c. Las funciones trigonométricas inversas—Multivalencia

Para formar las inversas de las funciones trigonométricas nuevamente se consideran las gráficas[1] de $\operatorname{sen} x$, $\cos x$, $\tan x$, y $\cot x$. Se ve de inmediato a partir de las Figs. 1.37, p. 74 y 1.38, p. 74, que para cada una de estas

[1] La representación gráfica ayudará al lector a superar las ligeras dificultades inherentes a la discusión de la "multivalencia" de las funciones inversas.

funciones es necesario seleccionar un intervalo definido si va a hablarse de una inversa única, pues las rectas $y = c$ paralelas al eje x cortan a las curvas en un número infinito de puntos.

El seno y el coseno inversos

Para la función $y = \operatorname{sen} x$ (Fig. 3.4), la derivada $y' = \cos x$ es positiva en el intervalo $-\pi/2 < x < \pi/2$. En este intervalo $y = \operatorname{sen} x$ posee una función inversa que denotamos por [1]

$$x = \operatorname{arc\,sen} y$$

(léase arco seno y; esto significa el ángulo cuyo seno tiene el valor y). Esta función crece monótonamente de $-\pi/2$ a $+\pi/2$ conforme y recorre el intervalo de -1 a $+1$. Si se desea destacar que estamos considerando la fun-

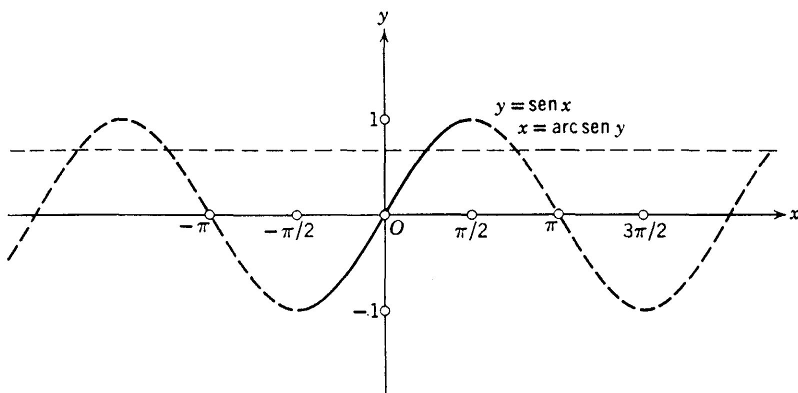


Figura 3.4 Gráfica de $y = \operatorname{sen} x$ (el valor principal está indicado por una curva llena).

ción inversa del seno en este intervalo particular, se habla del *valor principal* del arco seno. Para algún otro intervalo en el cual $\operatorname{sen} x$ es monótona, por ejemplo, el intervalo $+\pi/2 < x < 3\pi/2$, se obtiene otra inversa o *"rama"* del arco seno. Sin la especificación exacta del intervalo en el cual los valores de la función inversa deban estar situados, el símbolo arco seno no representa una función bien definida sino que, de hecho, denota un número *infinito* de valores.[2]

La multivalencia de $\operatorname{arc\,sen} y$ está descripta por la propiedad: A cualquier valor y del seno le corresponde no sólo un ángulo x específico sino también cualquier ángulo de la forma $2k\pi + x$, o bien $(2k + 1)\pi - x$, donde k es cualquier entero (véase Fig. 3.4).

[1] La notación simbólica $x = sen^{-1}\, y$ es utilizada también cuando no existe peligro de confusión con la función *recíproca* $1/\operatorname{sen} x$.
[2] A veces llamada vagamente función *multivaluada*.

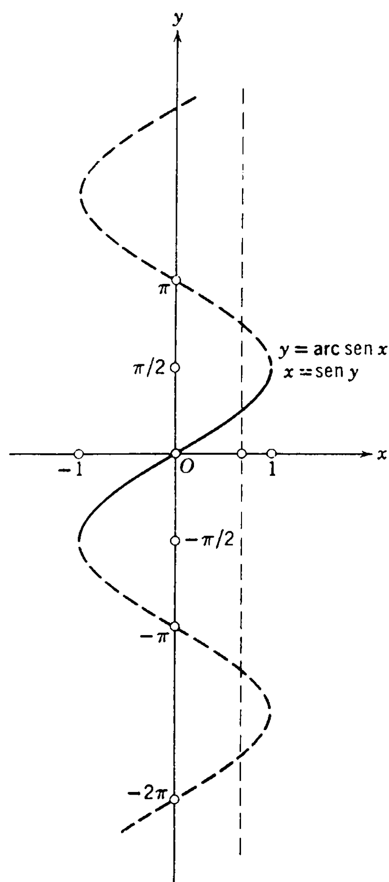


Figura 3.5 Gráfica de $y = \text{arc sen}\, x$ (el valor principal está indicado por una curva llena).

La derivada de la función $x = \text{arc sen}\, y$ se obtiene a partir de la Ec. (5) como sigue:

$$\frac{dx}{dy} = \frac{1}{y'} = \frac{1}{\cos x} = \frac{1}{\pm\sqrt{1 - \text{sen}^2 x}} = \frac{1}{\pm\sqrt{1 - y^2}},$$

donde la raíz cuadrada ha de tomarse como positiva si nos restringimos al primer intervalo mencionado, esto es, $-\pi/2 < x < \pi/2$.[1]

Finalmente, cambiamos el nombre de la variable independiente, de y al comúnmente usado x (Fig. 3.5); entonces la derivada de arc sen x está expresada por

---

[1] Si, en vez de éste, se hubiese elegido el intervalo $\pi/2 < x < 3\pi/2$, correspondiente a la substitución de x por $x + \pi$, se habría tenido que utilizar la raíz cuadrada negativa, puesto que $\cos x$ es negativo en este intervalo.

$$\frac{d}{dx}\,\text{arc sen}\,x = \frac{1}{\sqrt{1 - x^2}}.$$

Aquí se supone que arc sen es el valor principal que está situado entre $-\pi/2$ y $+\pi/2$, y el signo de la raíz cuadrada es elegido positivo.

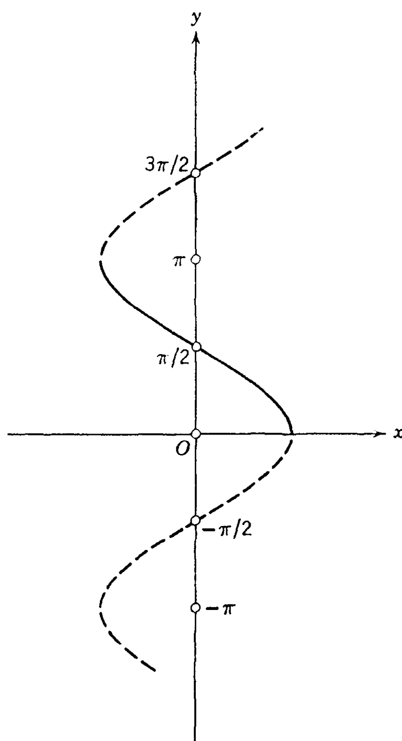


Figura 3.6 Gráfica de $y = \text{arc cos}\,x$ (el valor principal está indicado por una curva llena).

Para la función inversa de $y = \cos x$, denotada (después de intercambiar nuevamente los nombres x e y) por $\text{arc cos}\,x$, se obtiene la fórmula

$$\frac{d}{dx}\,\text{arc cos}\,x = \mp \frac{1}{\sqrt{1 - x^2}}$$

exactamente de la misma manera. Aquí se toma el signo negativo de la raíz si el valor de $\text{arc cos}\,x$ se toma en el intervalo entre 0 y π (no entre $-\pi/2$ y $+\pi/2$, como en el caso de $\text{arc sen}\,x$) (véase Fig. 3.6).

Las derivadas se vuelven infinitas al aproximarse a los puntos extremos $x = -1$ y $x = +1$, lo cual corresponde al hecho de que las gráficas del seno inverso y del coseno inverso tienen tangentes verticales en estos puntos.

Tangente y cotangente inversas

Las funciones inversas de la tangente y de la cotangente se tratan en forma análoga. La función $y = \tan x$, que posee en todas partes una derivada positiva $1/\cos^2 x$ para $x \neq \pi/2 + k\pi$, posee una inversa única en el intervalo $-\pi/2 < x < \pi/2$. Denominamos a esta función inversa (la Rama Principal de) $x = \operatorname{arc\,tan} y$. Se ve de inmediato a partir de la Fig. 3.7 que para cada x pudo haberse elegido en vez de y cualesquiera de los

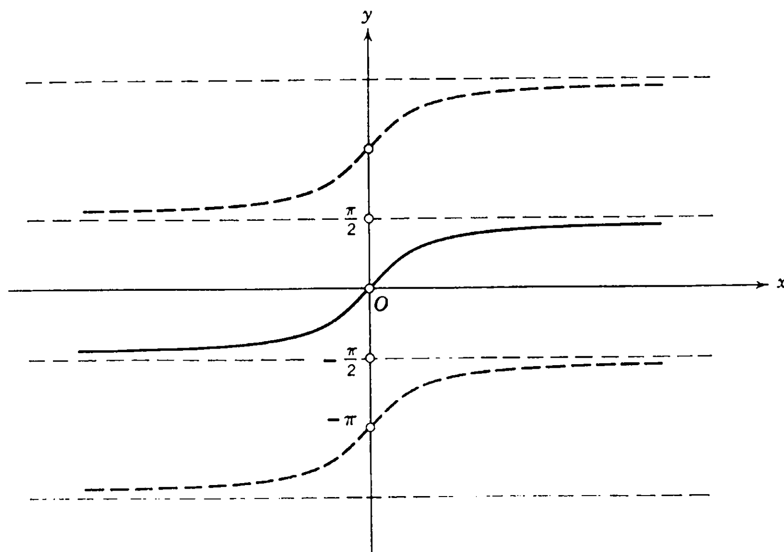


Figura 3.7 Gráfica de $y = \operatorname{arc\,tan} x$ (la curva llena para el valor principal).

valores $y + k\pi$ (donde k es un entero). Análogamente, la función $y = \cot x$ posee una inversa $x = \operatorname{arc\,cot} y$ que está determinada de manera única si se exige que su valor esté situado en el intervalo de 0 a π; si no, la multivalencia de $\operatorname{arc\,cot} x$ es la misma que para $\operatorname{arc\,tan} x$.

Las fórmulas de derivación son como sigue:

$$x = \operatorname{arc\,tan} y, \qquad \frac{dx}{dy} = \frac{1}{dy/dx} = \cos^2 x = \frac{1}{1 + \tan^2 x} = \frac{1}{1 + y^2},$$

$$x = \operatorname{arc\,cot} y, \qquad \frac{dx}{dy} = -\operatorname{sen}^2 x = -\frac{1}{1 + \cot^2 x} = -\frac{1}{1 + y^2},$$

o, finalmente, si denotamos nuevamente la variable independiente por x,

$$\frac{d}{dx} \operatorname{arc\,tan} x = \frac{1}{1 + x^2},$$

$$\frac{d}{dx} \operatorname{arc\,cot} x = -\frac{1}{1 + x^2}.$$

d. Las fórmulas integrales correspondientes

Expresadas en términos de integrales indefinidas, las fórmulas que se acaban de deducir se escriben como sigue:

$$\int \frac{1}{\sqrt{1-x^2}}\,dx = \text{arc sen}\, x, \qquad \int \frac{1}{\sqrt{1-x^2}}\,dx = -\text{arc cos}\, x,$$

$$\int \frac{1}{1+x^2}\,dx = \text{arc tan}\, x, \qquad \int \frac{1}{1+x^2}\,dx = -\text{arc cot}\, x.$$

Aún cuando las dos fórmulas de cada fila expresan diferentes funciones mediante integrales indefinidas idénticas, no hay contradicción entre ellas. En efecto, ilustran lo que se aprendió anteriormente (ver sección 2.9), a saber: que todas las integrales indefinidas de la misma función difieren solamente en constantes. Aquí las constantes son $\pi/2$ puesto que $\text{arc cos}\, x + \text{arc sen}\, x = \pi/2$, $\text{arc tan}\, x + \text{arc cot}\, x = \pi/2$.

Las fórmulas para integrales indefinidas pueden aplicarse de inmediato para encontrar integrales definidas, como en la p. 165. En particular,

$$\int_a^b \frac{dx}{1+x^2} = \text{arc sen}\, x \Big|_a^b = \text{arc tan}\, b - \text{arc tan}\, a.$$

Si se hacen $a = 0$, $b = 1$ y se recuerda que $\tan 0 = 0$ y $\tan \pi/4 = 1$, se obtiene la importante fórmula

$$\text{(6)} \qquad \frac{\pi}{4} = \int_0^1 \frac{1}{1+x^2}\,dx.$$

El número π, que surgió originalmente de la consideración del círculo, es puesto por esta fórmula en una relación muy simple con la función racional $1/(1+x^2)$ y representa el área indicada en la Fig. 3.8. Esta fórmula para π, a la cual regresaremos posteriormente (Sec. 5.2*a*), constituye uno de los primeros triunfos de la potencia del cálculo.

Más generalmente, las fórmulas integrales de esta sección nos permiten definir las funciones trigonométricas de manera puramente analítica, sin referencia alguna a objetos geométricos tales como triángulos o círculos. Por ejemplo, la relación entre el ángulo y y su tangente, $x = \tan y$, está descrita completamente por la ecuación

$$y = \int_0^x \frac{du}{1+u^2}$$

(al menos para $-\pi/2 < y < \pi/2$). Con esta relación podemos *definir* ahora, sin recurrir a la intuición, un valor numérico para el ángulo y en un triángulo rectángulo con lados a (adyacente) y b (opuesto) para el cual

$b/a = x$. Tal definición analítica en términos de cantidades numéricas hace lícito el uso de ángulos y funciones trigonométricas en el análisis superior independientemente de una definición por construcción geométrica.

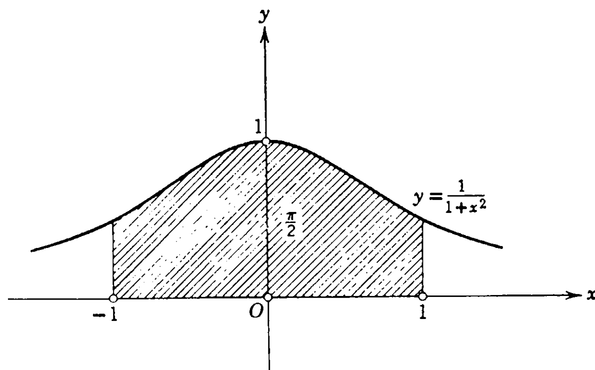


Figura 3.8 $\pi/2$ ilustrado mediante un área.

e. Derivada e integral de la función exponencial

En el capítulo 2, p. 172, se introdujo la función exponencial como la inversa del logaritmo. Más precisamente, las relaciones $y = e^x$ y $x = \log y$ fueron así definidas como equivalentes. En consecuencia, sus derivadas satisfacen la relación [ver (5), p. 229]

$$\frac{de^x}{dx} = \frac{dy}{dx} = \frac{1}{\dfrac{dx}{dy}} = \frac{1}{\dfrac{d\log y}{dy}} = \frac{1}{\dfrac{1}{y}} = y = e^x.$$

Por tanto, *la función exponencial es igual a su derivada:*

$$\frac{de^x}{dx} = e^x.$$

Más generalmente, para cualquier a positivo la función $y = a^x$ posee como inversa

$$x = \log_a y = \frac{\log y}{\log a},$$

y la derivada de a^x es

$$\frac{da^x}{dx} = \frac{1}{\dfrac{d\log_a y}{dy}} = (\log a)y = (\log a)a^x.$$

Así, para cualquier constante positiva a la derivada de la función $y = a^x$ es proporcional a la función misma. El factor de proporcionalidad $\log a$ es

1 cuando a es el número e. En la p. 244 se mostrará, recíprocamente que cualquier función que es proporcional a su derivada deberá ser de la forma $y = ce^x$, donde c denota un factor constante.

Por el teorema fundamental del cálculo, como antes, pueden traducirse las fórmulas para las derivadas de e^x y de a^x en fórmulas para integrales indefinidas:

$$\int e^x \, dx = e^x,$$

$$\int a^x \, dx = \frac{1}{\log a} a^x.$$

3.3 Derivación de funciones compuestas

a. Definiciones

Las reglas anteriores nos permiten encontrar las derivadas de funciones que se obtienen como expresiones racionales, en términos de funciones derivadas ya conocidas. Para encontrar expresiones explícitas para las derivadas de otras funciones que ocurren en análisis debemos llegar más lejos deduciendo una regla general para derivar funciones compuestas. Bastante a menudo se encuentran funciones $f(x)$ construidas por el proceso de composición de otras más sencillas (ver capítulo 1, p. 76): $f(x) = g(\phi(x))$, donde $\phi(x)$ está definida en un intervalo cerrado $a \leq x \leq b$ y tiene ahí el rango $\alpha \leq \phi \leq \beta$, y donde $g(\phi)$ está definida en este último intervalo.

En relación con esto es útil recordar la interpretación de funciones como "operadores" o transformaciones. Como en el capítulo 1, se escribe la función compuesta simplemente como

$$f = g\phi$$

y se denomina a $g\phi$ el "producto" (simbólico) de los operadores o transformaciones g y ϕ.

b. La regla de la cadena

Para funciones g y ϕ que son continuas en sus respectivos intervalos de definición, la función compuesta $f(x) = g[\phi(x)]$ es también continua (ver capítulo 1, p. 79).

Las funciones $\phi(x)$ y $g(\phi)$ se suponen ahora no sólo continuas sino también derivables. Se tiene entonces el siguiente teorema fundamental, de la *"regla de la cadena" para la derivación:*

La función $f(x) = g[\phi(x)]$ *es derivable y su derivada está dada por la ecuación*

$$f'(x) = g'(\phi) \cdot \phi'(x), \tag{7}$$

o bien, en la notación de Leibnitz,

$$\frac{df}{dx} = \frac{df}{d\phi}\frac{d\phi}{dx}.$$

Por consiguiente, la derivada de la función compuesta es el producto de las derivadas de las funciones que la constituyen. O bien: La derivada del producto simbólico de funciones es el producto real de sus derivadas con respecto a sus correspondientes variables independientes.

Intuitivamente esta regla de la cadena es plausible. La cantidad $\phi'(x) = \text{lím}\, \Delta\phi/\Delta x$ es el cociente local en el cual pequeños intervalos son amplificados por el "mapeo" ϕ. Análogamente, $g'(\phi)$ es la amplificación dada por el "mapeo" g. Aplicar primero ϕ y después g equivale a amplificar un intervalo x primero ϕ' veces y luego agrandar el intervalo ϕ resultante g' veces, obteniéndose así un cociente total de amplificación igual a $g'\phi'$, que debe ser el cociente de amplificación para el "mapeo" compuesto $f = g\phi$.

El teorema se sigue muy fácilmente a partir de la definición de la derivada. De hecho, se vuelve intuitivamente casi obvio si se supone $\phi'(x) \neq 0$ en el intervalo cerrado x que se considera. Entonces, para $\Delta x = x_2 - x_1 \neq 0$ se tiene por el teorema del valor medio

$$\Delta\phi = \phi_2 - \phi_1 = \phi(x_2) - \phi(x_1) = \phi'(\xi)\Delta x \neq 0 \quad \text{con} \quad x_1 \leq \xi \leq x_2,$$

y, con $\Delta g = g(\phi_2) - g(\phi_1)$ y $\Delta f = f(x_2) - f(x_1)$, puede escribirse

$$\frac{\Delta f}{\Delta x} = \frac{\Delta g}{\Delta\phi}\frac{\Delta\phi}{\Delta x}$$

que es una identidad con sentido, puesto que $\Delta\phi \neq 0$. Ahora $\Delta\phi \to 0$ para $\Delta x \to 0$, esto es, para $x_2 \to x_1$; por tanto, para $\Delta x \to 0$ los cocientes incrementales tienden a sus respectivas derivadas y el teorema está probado.

Para evitar la hipótesis explícita $\phi'(x) \neq 0$ podemos prescindir de la división por $\Delta\phi$ de la siguiente manera un poco más ingeniosa:

Por la hipótesis de la derivabilidad de $g(\phi)$ en el punto ϕ se sabe que la cantidad $\varepsilon = \Delta g/\Delta\phi - g'(\phi)$, como función de $\Delta\phi$ para ϕ fijo y $\Delta\phi \neq 0$, posee el límite cero para $\Delta\phi \to 0$. Si se *define* $\varepsilon = 0$ para $\Delta\phi = 0$, sin restricción sobre $\Delta\phi$ se tiene

$$\Delta g = [g'(\phi) + \varepsilon]\,\Delta\phi.$$

Análogamente, para x fijo,

$$\Delta\phi = \phi(x + \Delta x) - \phi(x) = [\phi'(x) + \eta]\,\Delta x,$$

donde $\lim_{\Delta x \to 0} \eta = 0$. Entonces, para $\Delta x \neq 0$ y $\phi = \phi(x)$, es

$$\frac{\Delta g}{\Delta x} = [g'(\phi) + \varepsilon]\frac{\Delta\phi}{\Delta x} = [g'(\phi) + \varepsilon][\phi'(x) + \eta].$$

Para Δx tendiendo a cero a través de valores distintos de cero se tiene $\lim_{\Delta x \to 0} \Delta \phi = 0$ y por lo tanto $\lim_{\Delta x \to 0} \varepsilon = 0$, de modo que

$$\lim_{\Delta x \to 0} \frac{\Delta g}{\Delta x} = \lim_{\Delta x \to 0} [g'(\phi) + \varepsilon] \lim_{\Delta x \to 0} [\phi'(x) + \eta] = g'(\phi)\phi'(x),$$

lo que demuestra la regla de la cadena.

Por aplicación sucesiva esta regla puede extenderse inmediatamente a funciones que resultan de la *composición de más de dos funciones.* Si, por ejemplo,

$$y = g(u), \qquad u = \phi(v), \qquad v = \psi(x),$$

entonces $y = f(x) = g[\phi(\psi(x))]$ es una función compuesta de x; su derivada está dada por la regla

$$\frac{dy}{dx} = y' = g'(u)\phi'(v)\psi'(x) = \frac{dy}{du}\cdot\frac{du}{dv}\cdot\frac{dv}{dy};$$

relaciones similares son ciertas para funciones que están compuestas de un número arbitrario de funciones.

Derivadas superiores de una función compuesta $y = g[\phi(x)]$. Pueden encontrarse fácilmente mediante la aplicación repetida de la regla de la cadena y las reglas anteriores:

$$y' = \frac{dy}{d\phi}\frac{d\phi}{dx} = g'\phi',$$
$$y'' = g''\phi'^2 + g'\phi'',$$
$$y''' = g'''\phi'^3 + 3g''\phi'\phi'' + g'\phi'''.$$

Fórmulas análogas para y'''', etc., pueden ser deducidas sucesivamente.

Finalmente, examinemos la *composición de dos funciones que son inversas* una con respecto a la otra. La función $g(y)$ es la inversa de $y = \phi(x)$ si $f(x) = g[\phi(x)] = x$. Se sigue que

$$f'(x) = g'(y)\phi'(x) = 1$$

que es exactamente el resultado de la sección 3.2, p. 228.

Ejemplos. Como un ejemplo sencillo, pero importante, relativo a una aplicación de la regla de la cadena derivemos $x^\alpha (x > 0)$ para una potencia real arbitraria α. En el capítulo 2, p. 174, se definió

$$x^\alpha = e^{\alpha \log x};$$

y se probó también para $\phi(x) = \log x, \psi(u) = \alpha u, g(y) = e^y$ que

$$\phi'(x) = \frac{1}{x} \qquad \psi'(u) = \alpha, \qquad g'(y) = e^y.$$

Ahora x^α es la función composición $g\{\psi[\phi(x)]\}$. Aplicando la regla de la cadena se obtiene la fórmula general

$$\begin{aligned}\frac{d}{dx}(x^\alpha) &= g'(y)\psi'(u)\phi'(x)\\ &= e^y\cdot\alpha\cdot\frac{1}{x}\\ &= \frac{\alpha e^{\alpha\log x}}{x}\\ &= \alpha\frac{x^\alpha}{x};\end{aligned}$$

por tanto

$$\frac{d}{dx}(x^\alpha) = \alpha x^{\alpha-1},$$

un resultado que solamente se podría probar con alguna dificultad si se intentara proceder directamente a partir de la definición de x^α para α irracional como el límite de potencias con exponentes racionales.

Una consecuencia inmediata de esta derivación es, nuevamente, la fórmula integral

$$\int x^\alpha\,dx = \frac{x^{\alpha+1}}{\alpha+1} \qquad (\alpha \neq -1).$$

Como un segundo ejemplo, considérese

$$y = \sqrt{1-x^2}, \qquad \text{o sea} \quad y = \sqrt{\phi},$$

donde $\phi = 1 - x^2$ con $-1 < x < 1$. La regla de la cadena conduce a

$$y' = \frac{1}{2\sqrt{\phi}}\cdot(-2x) = \frac{x}{\sqrt{1-x^2}}.$$

Los siguientes cálculos breves proporcionan ejemplos adicionales.

1. $y = \text{arc sen}\sqrt{1-x^2}, \qquad (-1 \leq x \leq 1, x \neq 0).$

$$\begin{aligned}\frac{dy}{dx} &= \frac{1}{\sqrt{1-(1-x^2)}}\frac{d\sqrt{1-x^2}}{dx}\\ &= \frac{1}{|x|}\frac{-x}{\sqrt{1-x^2}} = \frac{-1}{\sqrt{1-x^2}}\,\text{sgn}\,(x).\end{aligned}$$

2. $y = \sqrt{\dfrac{1+x}{1-x}}, \qquad (-1 < x < 1).$

$$\begin{aligned}\frac{dy}{dx} &= \frac{1}{2\sqrt{\dfrac{1+x}{1-x}}}\cdot\frac{d\left(\dfrac{1+x}{1-x}\right)}{dx}\\ &= \frac{\sqrt{1-x}}{2\sqrt{1+x}}\cdot\frac{2}{(1-x)^2} = \frac{1}{(1+x)^{1/2}(1-x)^{3/2}}.\end{aligned}$$

3. $y = \log |x|$. Esta función[1] puede expresarse como $\log x$ para $x > 0$ y como $\log(-x)$ para $x < 0$. Para $x > 0$ es

$$\frac{d \log |x|}{dx} = \frac{d \log x}{dx} = \frac{1}{x}.$$

Para $x < 0$, se obtiene a partir de la regla de la cadena que

$$\frac{d \log |x|}{dx} = \frac{d \log(-x)}{dx} = \frac{1}{-x} \frac{d(-x)}{dx} = \frac{1}{x}$$

Por tanto, en forma general para $x \neq 0$ es

$$\frac{d \log |x|}{dx} = \frac{1}{x}$$

4. $y = a^x$. Por definición de a^x (ver p. 174), se tiene

$$a^x = e^{\phi(x)},$$

donde $\phi(x) = (\log a)x$. Entonces,

$$\frac{da^x}{dx} = \frac{de^\phi}{d\phi} \frac{d\phi}{dx} = e^\phi(\log a) = (\log a)a^x.$$

El mismo resultado fue obtenido ya en la p. 238 a partir de la regla para la derivada de la función inversa.

5. $y = [f(x)]^{g(x)}$. Puesto que

$$[f(x)]^{g(x)} = e^{\phi(x)},$$

con $\phi(x) = g(x) \log [f(x)]$, se encuentra que

$$\frac{d}{dx}[f(x)]^{g(x)} = e^\phi \left(g' \log f + g \frac{1}{f} f' \right)$$

$$= [f(x)]^{g(x)} \left(g'(x) \log [f(x)] + \frac{g(x)f'(x)}{f(x)} \right).$$

Por ejemplo, cuando $g(x) = f(x) = x$ se tiene

$$\frac{dx^x}{dx} = x^x(\log x + 1).$$

c. El teorema generalizado del valor medio del cálculo diferencial

Como una aplicación de la regla de la cadena se deduce el *teorema generalizado del valor medio* del cálculo diferencial. Considérense dos funciones $F(x)$ y $G(x)$ continuas en un intervalo cerrado $[a, b]$ del eje x y derivables en el interior de este

---

[1] La función $\log x$ está definida solamente para $x > 0$, mientras que $\log |x|$ está definida en todas partes excepto para $x = 0$.

intervalo. Se supone $G'(x)$ positiva. El teorema ordinario del valor medio del cálculo diferencial aplicado separadamente a F y a G proporciona una expresión para el cociente de incrementos $\dfrac{F(b) - F(a)}{G(b) - G(a)}$:

$$\frac{F(b) - F(a)}{G(b) - G(a)} = \frac{F'(\xi)(b-a)}{G'(\eta)(b-a)} = \frac{F'(\xi)}{G'(\eta)}$$

donde ξ y η son valores intermedios apropiados en el intervalo abierto (a, b). El teorema generalizado del valor medio establece que el cociente de incrementos puede ser escrito en la forma más sencilla

$$\frac{F(b) - F(a)}{G(b) - G(a)} = \frac{F'(\zeta)}{G'(\zeta)}$$

donde F' y G' están evaluadas en *el mismo valor intermedio* ζ.

Para la demostración introducimos $u = G(x)$ como una variable independiente en F. De la hipótesis $G' > 0$ se concluye que la función $u = G(x)$ es monótona en el intervalo $[a, b]$, y, por tanto, que posee una inversa $x = g(u)$ definida en el intervalo $[\alpha, \beta]$, con $\alpha = G(a)$, $\beta = G(b)$. La función compuesta $F[g(u)] = f(u)$ está, por tanto, definida para u en el intervalo $[\alpha, \beta]$. A partir del teorema ordinario del valor medio se encuentra que

$$F(b) - F(a) = f(\beta) - f(\alpha) = f'(\gamma)(\beta - \alpha) = f'(\gamma)[G(b) - G(a)],$$

donde γ es un valor apropiado entre α y β. Por la regla de la cadena se infiere que

$$f'(u) = \frac{d}{du}F[g(u)] = F'[g(u)]g'(u) = \frac{F'(x)}{G'(x)}.$$

Al valor $u = \gamma$ le corresponde un valor $x = g(\gamma) = \zeta$ en el intervalo (a, b). Entonces $f'(\gamma) = F'(\zeta)/G'(\zeta)$, y se sigue el teorema generalizado del valor medio.

3.4 Algunas aplicaciones de la función exponencial

Algunos problemas diversos que involucran la función exponencial ilustrarán la importancia fundamental de esta función en toda clase de aplicaciones.

a. Definición de la función exponencial por medio de una ecuación diferencial

Podemos definir la función exponencial mediante una propiedad sencilla cuyo uso evita muchos argumentos detallados en casos particulares.

Si una función $y = f(x)$ satisface una ecuación de la forma

$$y' = \alpha y,$$

donde α es una constante, entonces y tiene la expresión

$$y = f(x) = ce^{\alpha x}, \tag{8}$$

donde c es también una constante; recíprocamente, toda función de la forma $ce^{\alpha x}$ satisface la ecuación $y' = \alpha y$.

Puesto que la Ec. (8) expresa una relación entre la función y su derivada, aquélla se denomina la *ecuación diferencial* de la función exponencial.

Es claro que $y = ce^{\alpha x}$ satisface esta ecuación para cualquier constante arbitraria c. Recíprocamente, ninguna otra función satisface la ecuación diferencial $y' - \alpha y = 0$. En efecto, si y es tal función, considérese la función $u = ye^{-\alpha x}$. Se tiene entonces

$$u' = y'e^{-\alpha x} - \alpha ye^{-\alpha x} = e^{-\alpha x}(y' - \alpha y).$$

Sin embargo, el miembro derecho se anula, ya que se ha supuesto que $y' = \alpha y$; por lo tanto $u' = 0$, de modo que (ver p. 199) u es una constante c, y así resulta $y = ce^{\alpha x}$, como se quería probar.

Aplicaremos ahora este teorema a cierto número de ejemplos.

b. Interés compuesto continuamente
Desintegración radiactiva

Una suma "capital" o "principal", aumentada por su interés en períodos regulares de tiempo, se incrementa por saltos en estos períodos de la siguiente manera. Si 100α es el porcentaje de interés y, además, si el interés acumulado se suma, al finalizar cada año, al capital invertido (original), después de x años la cantidad acumulada de un capital original igual a 1 será

$$(1 + \alpha)^x.$$

Sin embargo, si al capital le fuera sumado el interés no al finalizar cada año sino al finalizar cada n-ésima parte de un año, después de x años el capital ascendería a

$$\left(1 + \frac{\alpha}{n}\right)^{nx}.$$

Tomando, por simplicidad, $x = 1$ se encuentra que el capital 1 ha aumentado después de un año a

$$\left(1 + \frac{\alpha}{n}\right)^{n}.$$

Si ahora se deja crecer n más allá de toda cota, esto es, si se deja que el interés sea abonado en intervalos cada vez más cortos, el caso límite significará en cierto sentido que el interés compuesto es abonado a cada instante; entonces la cantidad total después de un año será e^{α} veces el capital inicial u original (véase p. 175). Análogamente, si el interés se calcula de esta manera, un capital original igual a 1 habrá crecido después de x años hasta una cantidad $e^{\alpha x}$; aquí x puede ser cualquier número, entero o de otra forma.

La discusión de la sección 3.4a constituye un marco de referencia en el cual ejemplos de este tipo encajan fácilmente. Considérese una cantidad, dada por el número y, que crece (o decrece) con el tiempo de modo que la medida en que la cantidad aumenta o disminuye es proporcional a la cantidad total. Entonces, tomando el tiempo como la variable independiente x se obtiene una ley de la forma $y' = \alpha y$ para el grado de aumento (o decrecimiento), donde α, el factor de proporcionalidad, es positivo o negativo dependiendo de que la cantidad esté creciendo o decreciendo. Entonces, de acuerdo con la sección 3.4a, la cantidad y está representada por la fórmula

$$y = ce^{\alpha x},$$

donde el significado de la constante c es obvio si se considera el instante $x = 0$. En este instante $e^{\alpha x} = 1$, y se encuentra que $c = y_0$ es la cantidad al comienzo del tiempo considerado, de modo que se puede escribir

$$y = y_0 e^{\alpha x}.$$

Un ejemplo característico es el de la *desintegración radiactiva.* La medida en que la cantidad total y de la substancia radiactiva está disminuyendo es proporcional en cualquier instante a la cantidad total presente en ese instante; esto es plausible *a priori* pues cada porción de la substancia decrece tan rápidamente como cualquier otra porción. Por tanto, la cantidad y de la substancia, expresada como función del tiempo, satisface una relación de la forma $y = -ky$, donde k debe tomarse positiva puesto que se está tratando con una cantidad que disminuye. La cantidad de substancia queda así expresada como función del tiempo por $y = y_0 e^{-kx}$, donde y_0 es la cantidad de la substancia al comienzo del tiempo considerado (tiempo $x = 0$).

Después de un cierto tiempo τ la substancia radiactiva habrá disminuido a la mitad de la cantidad original. Este tiempo, llamado *vida media,* está dado por la ecuación

$$\tfrac{1}{2}y_0 = y_0 e^{-k\tau},$$

de la cual se obtiene de inmediato $\tau = (\log 2)/k$.

c. Enfriamiento o calentamiento de un cuerpo por el medio ambiente

Otro ejemplo típico en el que aparece la función exponencial lo constituye el enfriamiento de un cuerpo; por ejemplo, una placa de metal de temperatura uniforme que es sumergida en un gran "baño", o medio circundante o "ambiente", de baja temperatura. Se supone que la extensión del medio ambiente es tan grande que su temperatura no es afectada por el proceso de enfriamiento. Se supone además que en cada instante todas las partes del cuerpo sumergido se encuentran a la misma temperatura, y

que la medida en que la temperatura varía es proporcional a la diferencia entre la temperatura del cuerpo y la del medio circundante (Ley de enfriamiento de Newton).

Si se denota el tiempo por x y la diferencia de temperatura entre el cuerpo y el baño por $y = y(x)$, esta ley de enfriamiento está expresada por la ecuación

$$y' = -ky,$$

donde k es una constante positiva (cuyo valor es una característica física de la substancia del cuerpo). A partir de esta ecuación diferencial, que expresa el efecto del proceso de enfriamiento en un instante dado, se obtiene por medio de la Ec. (8), p. 244, una "ley integral" que nos da la temperatura en un instante arbitrario x en la forma

$$y = ce^{-kx}.$$

Esto muestra que la temperatura decrece "exponencialmente" y tiende a ser igual a la temperatura externa. La rapidez con la que esto sucede está expresada por el número k. Como antes, el significado de la constante c es el de la temperatura inicial en el instante $x = 0$, $y_0 = c$, de modo que nuestra ley de enfriamiento puede ser escrita en la forma

$$y = y_0e^{-kx}.$$

Evidentemente, la misma discusión se aplica al calentamiento de un cuerpo. Lo único que cambia es que la diferencia inicial de temperaturas y_0 es en este caso negativa en vez de positiva.

d. Variación de la presión atmosférica con la altura sobre la superficie de la Tierra

Un ejemplo más en el que aparece la fórmula exponencial es el de la variación de la presión atmosférica con la altura: Se hará uso (1) del hecho físico de que la presión atmosférica es igual al peso de la columna de aire que está verticalmente por encima de una superficie de área uno, y (2) de la Ley de Boyle, de acuerdo con la cual la presión p del aire a una temperatura constante es proporcional a su densidad σ. La Ley de Boyle, expresada en símbolos, es $p = a\sigma$, donde a es una constante que depende de una propiedad física específica del aire. Nuestro problema es el de determinar $p = f(h)$ como una función de la altura h por encima de la superficie de la Tierra.

Si por p_0 se denota la presión atmosférica en la superficie de la Tierra, esto es, el peso total de la columna de aire sostenida por un área unitaria, por g la constante gravitacional y por $\sigma(\lambda)$ la densidad del aire a la altu-

ra λ sobre la superficie terrestre, el peso [1] de la columna de altura h está dado por la integral $g\int_0^h \sigma(\lambda)d\lambda$. La presión a una altura h es, por lo tanto,

$$p = f(h) = p_0 - g\int_0^h \sigma(\lambda)\, d\lambda.$$

Derivando esta expresión se obtiene la siguiente relación entre la presión $p = f(h)$ y la densidad $\sigma(h)$:

$$g\sigma(h) = -f'(h) = -p'.$$

Utilizando ahora la Ley de Boyle para eliminar la cantidad σ de esta ecuación, resulta una relación $p' = -(g/a)p$ que involucra solamente la función presión desconocida. De la Ec. (8), p. 244, se sigue que

$$p = f(h) = ce^{-gh/a}.$$

Si, como antes, la presión $f(0)$ en la superficie de la Tierra se denota por p_0, se sigue de inmediato que $c = p_0$, y consecuentemente

$$p = f(h) = p_0 e^{-gh/a}.$$

Tomando logaritmos esto conduce a

$$h = \frac{a}{g}\log\frac{p_0}{p}.$$

Estas dos fórmulas se aplican frecuentemente. Por ejemplo, la constante *a,* si es conocida, permite determinar la altura de un lugar a partir de la presión barométrica, o bien determinar la diferencia entre las alturas de dos lugares midiendo la presión atmosférica en cada sitio. Por otra parte, si la presión atmosférica y la altura h son conocidas, puede determinarse la constante *a,* la cual es de gran importancia en la teoría de los gases.

e. Progreso de una reacción química

Se considerará ahora un ejemplo de la Química, la así llamada *reacción unimolecular.* Se supone que una substancia es disuelta en una gran cantidad de solvente; por ejemplo, una cierta cantidad de azúcar de caña en agua. Si ocurre una reacción química, la ley química de acción de las masas establece en este caso que la medida de reacción es proporcional a la cantidad de substancia reaccionante presente. Se supone que el azúcar de caña está siendo transformada mediante acción catalítica en azúcar invertida, y se denota por $u(x)$ la cantidad de azúcar de caña que al tiempo x permanece invariable aún. La velocidad de reacción es entonces $-du/dx$,

[1] $g\sigma(\lambda)$ es el peso del aire por unidad de volumen a la altura λ.

y, de acuerdo con la ley de acción de masas, se cumple una ecuación de la forma

$$\frac{du}{dx} = -ku$$

donde k es una constante que depende de las substancias reaccionantes. A partir de esta ley instantánea o diferencial se obtiene de inmediato, como en la p. 244, una ley integral que da la cantidad de azúcar de caña como función del tiempo:

$$u(x) = ae^{-kx}.$$

Esta fórmula muestra claramente cómo la reacción química tiende asintóticamente a su estado final $u = 0$, esto es, la transformación completa de la substancia reaccionante. La constante a es obviamente la cantidad de azúcar de caña presente en el tiempo $x = 0$.

f. Apertura y cierre de un circuito eléctrico

Como un ejemplo final, considérese el crecimiento de una corriente eléctrica directa o "continua" cuando se cierra un circuito, o bien su decaimiento cuando el circuito se abre. Si R es la resistencia del circuito y E la fuerza electromotriz (voltaje), la corriente I se incrementa gradualmente desde su valor original cero hasta un valor estacionario final E/R. Tenemos, por tanto, que considerar I como una función del tiempo x. El aumento de la corriente depende de la autoinducción del circuito; y el circuito posee una constante característica L, el *coeficiente de autoinducción*, de naturaleza tal que conforme la corriente crece se desarrolla una fuerza electromotriz de magnitud $L\,dI/dx$ que se opone a la fuerza electromotriz externa E. A partir de la ley de Ohm, que afirma que el producto de la resistencia y de la corriente es en cada instante igual al voltaje efectivo real, se obtiene la relación

$$IR = E - L\frac{dI}{dx}.$$

para

$$f(x) = I(x) - \frac{E}{R}$$

se encuentra de inmediato que $f'(x) = -(R/L)f(x)$, de modo que por la Ec. (8), p. 244, es $f(x) = f(0)e^{-Rx/L}$. Recordando que $I(0) = 0$, se encuentra que $f(0) = -E/R$; y así se obtiene la expresión

$$I = f(x) + \frac{E}{R} = \frac{E}{R}(1 - e^{-Rx/L})$$

para la corriente como una función del tiempo.

Esta expresión muestra cómo la corriente tiende asintóticamente a su valor estacionario E/R cuando se cierra el circuito.

3.5 Las funciones hiperbólicas

a. Definición analítica

En muchas aplicaciones la función exponencial aparece en combinaciones de la forma

$$\tfrac{1}{2}(e^x + e^{-x}) \qquad \text{o bien} \quad \tfrac{1}{2}(e^x - e^{-x}).$$

Es conveniente introducir éstas y combinaciones similares como funciones especiales, que se definen como sigue:

$$\text{(9a)} \qquad \operatorname{senh} x = \frac{e^x - e^{-x}}{2}, \qquad \cosh x = \frac{e^x + e^{-x}}{2},$$

$$\text{(9b)} \qquad \tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \qquad \coth x = \frac{e^x + e^{-x}}{e^x - e^{-x}},$$

llamadas *seno hiperbólico, coseno hiperbólico, tangente hiperbólica,* y *cotangente hiperbólica* respectivamente. Las funciones $\operatorname{senh} x$, $\cosh x$, y $\tanh x$ están definidas para todos los valores de x, mientras que para $\coth x$ el punto $x = 0$ debe ser excluido. Los nombres son elegidos para expresar una cierta analogía con las funciones trigonométricas; y esta analogía, que ha de estudiarse en detalle, es la que justifica una consideración especial de estas nuevas funciones. En las Figs. 3.9, 3.10 y 3.11 se muestran las gráficas de las funciones hiperbólicas; y las líneas punteadas en la Fig. 3.9 son las gráficas de $y = (\frac{1}{2})e^x$ y de $y = (\frac{1}{2})e^{-x}$, a partir de las cuales pueden construirse fácilmente las gráficas de $\operatorname{senh} x$ y $\cosh x$.

Cosh x es, evidentemente, una *función par,* esto es, una función que permanece invariable cuando x es reemplazado por $-x$, mientras que $\operatorname{senh} x$ es una *función impar,* esto es, una función que cambia de signo cuando x es reemplazado por $-x$ (ver p. 53).

Por su definición, la función

$$\cosh x = \frac{e^x + e^{-x}}{2}$$

es positiva y no menor que uno para todos los valores de x. Posee su valor mínimo cuando $x = 0$: $\cosh 0 = 1$.

La relación fundamental entre $\cosh x$ y $\operatorname{senh} x$,

$$\cosh^2 x - \operatorname{senh}^2 x = 1,$$

se sigue de inmediato a partir de las definiciones. Si se denota ahora la variable independiente por t, en vez de x, y se escribe

$$x = \cosh t, \qquad y = \operatorname{senh} t,$$

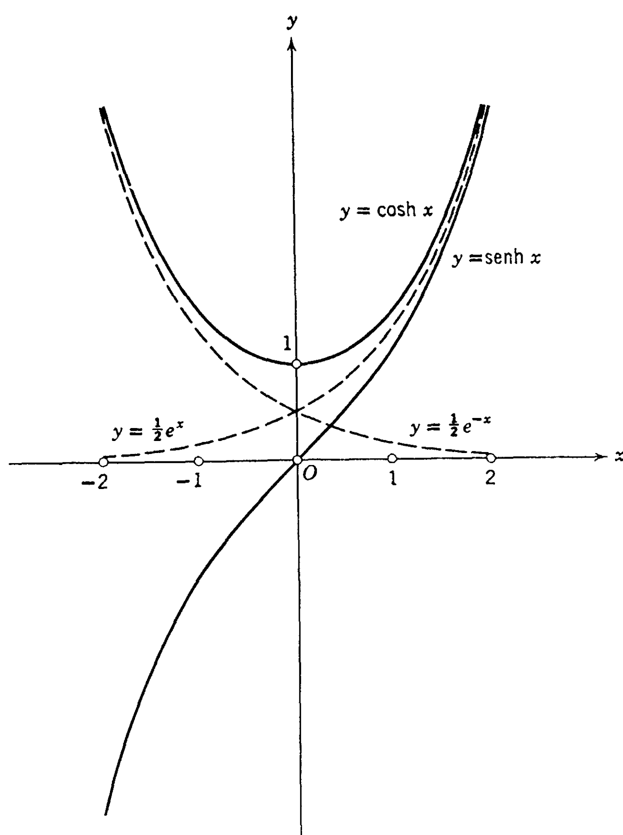


Figura 3.9.

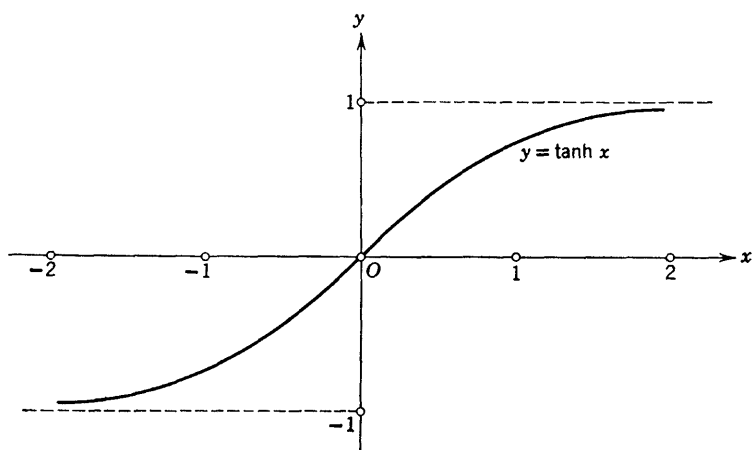


Figura 3.10.

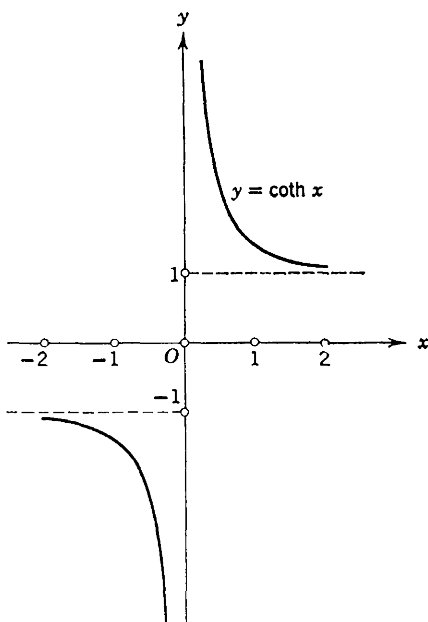


Figura 3.11.

se tiene

$$x^2 - y^2 = 1;$$

esto es, el punto de coordenadas $x = \cosh t$, $y = \operatorname{senh} t$ se mueve a lo largo de la hipérbola rectangular $x^2 - y^2 = 1$ conforme t recorre toda la escala de valores de $-\infty$ a $+\infty$. De acuerdo con la ecuación de definición, $x \geq 1$, y la fórmula hace evidente que y recorre toda la escala de valores de $-\infty$ a $+\infty$ conforme t lo hace; en efecto, si t tiende a infinito, también e^t, mientrsa que e^{-t} tiende a cero. Se puede por tanto establecer más exactamente: Conforme t corre de $-\infty$ a $+\infty$ las ecuaciones $x = \cosh t$, $y = \operatorname{senh} t$ dan una rama de la hipérbola rectangular, a saber, la del lado derecho.

b. Teoremas de adición y fórmulas para derivar

A partir de su definición se obtienen los *teoremas de adición* para las funciones hiperbólicas:

$$(10)\qquad \begin{aligned} \cosh(a + b) &= \cosh a \cosh b + \operatorname{senh} a \operatorname{senh} b, \\ \operatorname{senh}(a + b) &= \operatorname{senh} a \cosh b + \cosh a \operatorname{senh} b. \end{aligned}$$

Las demostraciones se obtienen de inmediato si se escribe

$$\cosh(a + b) = \frac{e^a e^b + e^{-a} e^{-b}}{2} \qquad \operatorname{senh}(a + b) = \frac{e^a e^b - e^{-a} e^{-b}}{2},$$

y se insertan en estas ecuaciones

$$e^a = \cosh a + \operatorname{senh} a, \qquad e^{-a} = \cosh a - \operatorname{senh} a,$$

$$e^b = \cosh b + \operatorname{senh} b, \qquad e^{-b} = \cosh b - \operatorname{senh} b.$$

Entre estas fórmulas y las correspondientes fórmulas trigonométricas existe una notable analogía. La única diferencia entre los teoremas de adición es un signo en la primera fórmula.

Una analogía correspondiente se cumple para las fórmulas de diferenciación. Recordando que $d(e^x)/dx = e^x$, se encuentra fácilmente que

$$\frac{d}{dx}\cosh x = \operatorname{senh} x, \qquad \frac{d}{dx}\operatorname{senh} x = \cosh x,$$

(11)

$$\frac{d}{dx}\tanh x = \frac{1}{\cosh^2 x}, \qquad \frac{d}{dx}\coth x = -\frac{1}{\operatorname{senh}^2 x}.$$

De las dos primeras ecuaciones se sigue de inmediato que $y = \cosh x$ e $y = \operatorname{senh} x$ son soluciones de la *ecuación diferencial*

(12)
$$\frac{d^2y}{dx^2} = y,$$

la cual difiere sólo en signo de la ecuación análoga que es satisfecha por las funciones trigonométricas $\cos x$ y $\operatorname{sen} x$ (véase p. 192).

c. Las funciones hiperbólicas inversas

A las funciones hiperbólicas $x = \cosh t$, $y = \operatorname{senh} t$, les corresponden las funciones inversas que denotamos[1] por

$$t = \operatorname{arc\,cosh} x, \qquad t = \operatorname{arc\,senh} y.$$

Puesto que la función $\operatorname{senh} t$ es monótona creciente[2] en todo el intervalo $-\infty < t < +\infty$, su función inversa está determinada de manera única para todos los valores de y; por otra parte, una ojeada a la gráfica (véase Fig. 3.9, p. 250) muestra que $t = \operatorname{arc\,cosh} x$ no está determinado de manera única sino que posee una ambigüedad de signo debido a que a un valor dado de x le corresponde no solamente el número t sino también el número $-t$. Puesto que $\cosh t \geq 1$ para todos los valores de t, su inversa $\operatorname{arc\,cosh} x$ está definida solamente para $x \geq 1$.

Estas funciones inversas pueden expresarse fácilmente en términos del logaritmo, interpretando la cantidad $e^t = u$ en las definiciones

$$x = \frac{e^t + e^{-t}}{2}, \qquad y = \frac{e^t - e^{-t}}{2}$$

---

[1] La notación *simbólica* $\cosh^{-1} x$, etc., es también utilizada (ver nota, página, 78).

[2] $(d/dt)\operatorname{senh} t = \cosh t > 0.$

como incógnita y resolviendo estas ecuaciones (cuadráticas) para u:

$$u = x \pm \sqrt{x^2 - 1}, \qquad u = y + \sqrt{y^2 + 1};$$

y puesto que $u = e^t$ sólo puede poseer valores positivos, la raíz cuadrada en la segunda ecuación debe tomarse con el signo positivo, mientras que en la primera cualquier signo es posible (lo que corresponde a la ambigüedad mencionada arriba). En la forma logarítmica, $t = \log u$, y por tanto

$$(13) \qquad \begin{aligned} t &= \log\,(x \pm \sqrt{x^2 - 1}) = \operatorname{arc\,cosh} x, \\ t &= \log\,(y + \sqrt{y^2 + 1}) = \operatorname{arc\,senh} y. \end{aligned}$$

En el caso de $\operatorname{arc\,cosh} x$ la variable x está restringida al intervalo $x \geq 1$, mientras que $\operatorname{arc\,senh} y$ está definido para todos los valores de y.

La ecuación (13) da los valores,

$$\log\,(x + \sqrt{x^2 - 1}) \qquad \text{y} \qquad \log\,(x - \sqrt{x^2 - 1})$$

para $\operatorname{arc\,cosh} x$, que corresponden a las dos ramas de $\operatorname{arc\,cosh} x$. Puesto que

$$(x + \sqrt{x^2 - 1})\,(x - \sqrt{x^2 - 1}) = 1,$$

la suma de estos dos valores de $\operatorname{arc\,cosh} x$ es cero, lo que concuerda con la ambigüedad en el signo de t mencionada antes.

Las inversas de la tangente hiperbólica y la cotangente hiperbólica pueden ser definidas análogamente, y también pueden ser expresadas en términos de logaritmos. Estas funciones las denotamos por $\operatorname{arc\,tanh} x$ y $\operatorname{arc\,coth} x$. Expresando la variable independiente en todas partes por x, fácilmente se obtiene

$$(14) \qquad \begin{aligned} \operatorname{arc\,tanh} x &= \frac{1}{2} \log \frac{1 + x}{1 - x} \qquad \text{en el intervalo } -1 < x < 1, \\ \operatorname{arc\,coth} x &= \frac{1}{2} \log \frac{x + 1}{x - 1} \qquad \text{en los intervalos } x < -1,\ x > 1. \end{aligned}$$

La diferenciación de estas funciones inversas puede ser llevada a cabo por el mismo lector, quien puede hacer uso ya sea de la regla para derivar una función inversa, o bien de la regla de la cadena simultáneamente con estas expresiones para las funciones inversas en términos de logaritmos. Si x es la variable independiente, los resultados son

$$(15) \qquad \begin{aligned} &\frac{d}{dx} \operatorname{arc\,cosh} x = \pm \frac{1}{\sqrt{x^2 - 1}}, \qquad &&\frac{d}{dx} \operatorname{arc\,senh} x = \frac{1}{\sqrt{x^2 + 1}}, \\ &\frac{d}{dx} \operatorname{arc\,tanh} x = \frac{1}{1 - x^2}, \qquad &&\frac{d}{dx} \operatorname{arc\,coth} x = \frac{1}{1 - x^2}. \end{aligned}$$

Las últimas dos fórmulas no se contradicen una con otra puesto que la primera se cumple solamente para $-1 < x < 1$ y la segunda solamente para $x < -1$ y $1 < x$. Los dos valores de la derivada $d(\text{arc cosh } x)/dx$, expresados por el signo $\pm$ en esta primera fórmula, corresponden a las dos ramas diferentes de la curva $y = \text{arc cosh } x = \log(x \pm \sqrt{x^2 - 1})$.

d. Otras analogías

Las semejanzas entre las funciones hiperbólicas y las trigonométricas no son accidentales. El origen profundo de estas analogías se manifiesta cuando se consideran estas funciones con argumentos imaginarios, como se hará posteriormente en la sección 7.7a. Se estará entonces en condiciones de identificar $\cosh x$ con $\cos(ix)$ y $\operatorname{senh} x$ con $(1/i)\operatorname{sen}(ix)$, donde $i = \sqrt{-1}$. Este hecho hace evidente que toda relación que involucra funciones trigonométricas tiene su contraparte para funciones hiperbólicas. Muchas de esas analogías poseen interesantes interpretaciones geométricas o físicas. (Véase también el capítulo 4, Sec. 4.1*j*.)

En la representación anterior de la hipérbola rectangular mediante la cantidad t no se atribuyó ningún significado geométrico a este "parámetro". Volviendo ahora a este tema se encuentra otra analogía entre las funciones trigonométricas y las hiperbólicas. Si la circunferencia cuya ecuación es $x^2 + y^2 = 1$ se representa por medio de un parámetro t en la forma $x = \cos t$, $y = \operatorname{sen} t$, la cantidad t puede interpretarse como un ángulo o bien como una longitud de arco medida a lo largo de la circunferencia. Podemos, sin embargo, interpretar t también como el doble del área del sector circular que corresponde a este ángulo, siendo considerada el área positiva o negativa dependiendo de si el ángulo es positivo o negativo.

Se establece ahora análogamente que para las funciones hiperbólicas la cantidad t es el doble del área del sector hiperbólico para $x^2 - y^2 = 1$, mostrado sombreado en la Fig. 3.12.[1] Es esta interpretación de t en términos de áreas la que explica y justifica los nombres $t = \text{arc cosh } x$ y $t = \text{arc senh } y$ dados a las funciones hiperbólicas inversas.[2] La demostración se obtiene sin ninguna dificultad si se refiere la hipérbola a sus asíntotas tomadas como ejes, por medio de la transformación de coordenadas

$$x - y = \sqrt{2}\xi, \qquad x + y = \sqrt{2}\eta,$$

o bien

$$x = \frac{1}{\sqrt{2}}(\xi + \eta), \qquad y = \frac{1}{\sqrt{2}}(\eta - \xi);$$

---

[1] Para una demostración diferente, véase Sec. 4.1*j*.

[2] Tal como la notación $t = \arccos x$ se refiere a un *arco* del círculo unitario, la notación $t = \text{arc cosh } x$ se refiere a un *área* conectada con una hipérbola rectangular $x^2 - y^2 = 1$. Debe notarse que t *no es* la longitud del arco hiperbólico.

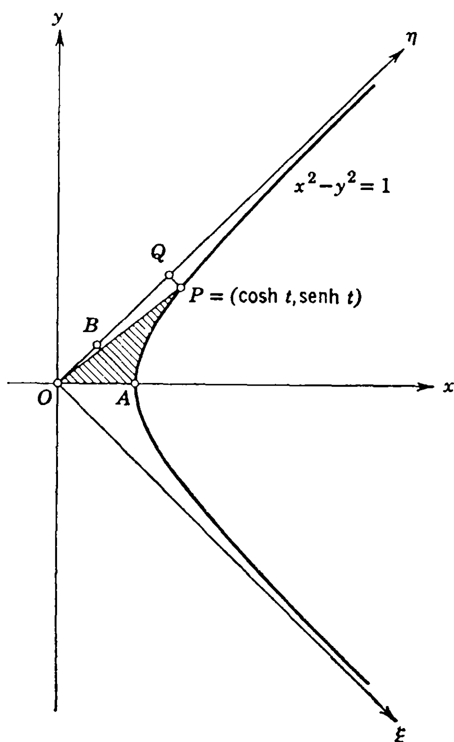


Figura 3.12.

con estas nuevas coordenadas la ecuación de la hipérbola es $\xi\eta = \frac{1}{2}$. Por tanto, los dos triángulos rectángulos OPQ y OAB poseen ambos un área igual a $\frac{1}{4}$, pues las longitudes de OQ y QP son, respectivamente, η y $1/2\eta$, y el área en cuestión es igual a la de la figura $ABQP$. Obviamente las coordenadas de los puntos A y B son

$$\xi = \frac{1}{\sqrt{2}}, \eta = \frac{1}{\sqrt{2}} \quad \text{y} \quad \xi = \frac{x-y}{\sqrt{2}}, \eta = \frac{x+y}{\sqrt{2}}$$

respectivamente, y para el doble del área de la figura en cuestión se obtiene así

$$2\int_{1/\sqrt{2}}^{(x+y)/\sqrt{2}} (1/2\eta)\, d\eta = \log(x+y) = \log(x \pm \sqrt{x^2-1}),$$

pero, por la Ec. (13), p. 254, esto es igual a t, lo que prueba la aseveración hecha.

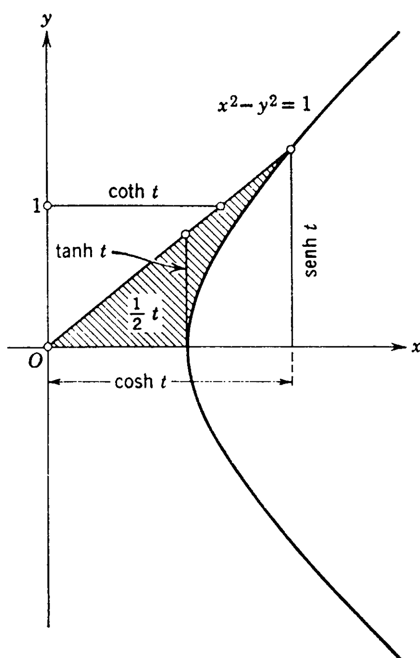


Figura 3.13 Para ilustrar las funciones hiperbólicas.

En conclusión, puede señalarse que, como se muestra en la Fig. 3.13, las funciones hiperbólicas pueden ser representadas gráficamente sobre la hipérbola, de la misma manera que las funciones trigonométricas pueden ser representadas sobre el círculo.

3.6 Máximos y mínimos

Como la primera de una gran variedad de aplicaciones se considerará ahora la teoría de máximos y mínimos de una función, junto con una discusión geométrica de la segunda derivada.

a. Convexidad y concavidad de curvas

Por definición, la derivada $f'(x) = df(x)/dx$ representa la pendiente de la curva $y = f(x)$. La derivada de la función $f'(x)$, es decir, de la pendiente de la curva $y = f(x)$, está dada por la derivada $df'(x)/dx = f''(x)$, la *segunda derivada* de $f(x)$; y así sucesivamente. Si la segunda derivada $f''(x)$ es positiva en un punto x—de modo que debido a la continuidad

(que se supone) es positiva en alguna vecindad [1] de este punto x—entonces a lo largo de esta vecindad la derivada $f'(x)$ aumenta con valores crecientes de x. Por lo tanto, la curva $y = f(x)$ tiene su lado convexo hacia abajo y es "abierta" hacia arriba. A la función $f(x)$, o a la curva $y = f(x)$, la denominamos *convexa*. Si $f''(x)$ es negativa, la curva y la función son *cóncavas*. Por lo tanto, cuando $f''(x) > 0$ la curva en la vecindad del punto

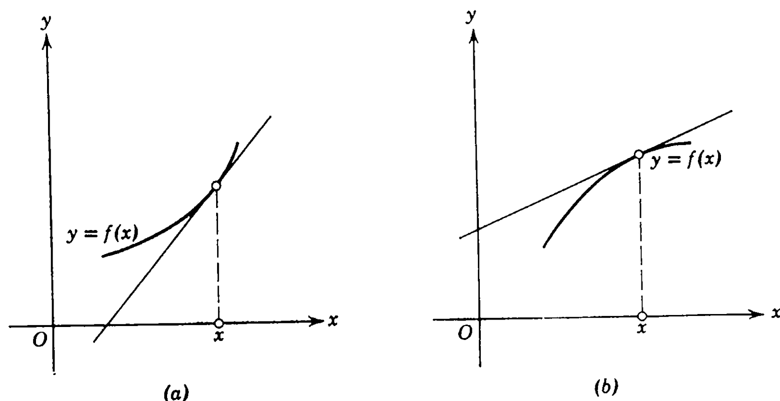


Figura 3.14 $(a) f''(x) > 0$. $(b) f''(x) < 0$.

está situada por encima de la tangente, mientras que cuando $f''(x) < 0$ está situada por debajo de la tangente (véanse las Figs. 3.14a y 3.14b) (ver problema 4, p. 221 y sección 5.6).

Punto de inflexión

Se requiere una consideración especial solamente en los puntos donde $f''(x) = 0$. Al pasar a través de uno de tales puntos la segunda derivada $f''(x)$ cambiará generalmente de signo. Tal punto será entonces un punto de transición entre los dos casos que se acaban de indicar; esto es, en un lado la tangente está por encima de la curva y por el otro lado está por debajo, mientras que en este punto cruza la curva (véase Fig. 3.15). Tal punto es denominado un *punto de inflexión* de la curva, y la tangente correspondiente se denomina *tangente de inflexión*.

[1] Se hace uso aquí de la observación intuitivamente obvia: una función *continua* $g(x)$ que es positiva en un punto x_0 es también positiva para todos los puntos de una vecindad suficientemente pequeña de x_0 (siempre que pertenezcan al dominio de g). La demostración rigurosa es sencilla. A partir de la continuidad de g en x_0, se sabe que para *todo* ε positivo se cumple la desigualdad $|g(x) - g(x_0)| < \varepsilon$ para todo x en una vecindad suficientemente pequeña $|x - x_0| < \delta$ del punto x_0. Como $g(x_0) > 0$, se tiene libertad de elegir para ε el valor $\frac{1}{2}g(x_0)$, de modo que $|g(x) - g(x_0)| < \frac{1}{2}g(x_0)$ en alguna vecindad. Puesto que entonces es $g(x_0) - g(x) \leq |g(x) - g(x_0)| < \frac{1}{2}g(x_0)$, se sigue que $g(x) > \frac{1}{2}g(x_0) > 0$.

El ejemplo más sencillo está dado por la función $y = x^3$, la parábola cúbica, para la cual el eje x mismo es una tangente de inflexión en el punto de inflexión $x = 0$ (véase Fig. 3.3, p. 231). Otro ejemplo está dado por la función $f(x) = \operatorname{sen} x$, para la cual

$$f'(x) = d(\operatorname{sen} x)/dx = \cos x \qquad \text{y} \qquad f''(x) = d^2(\operatorname{sen} x)/dx^2 = -\operatorname{sen} x.$$

Consecuentemente, $f'(0) = 1$ y $f''(0) = 0$; y, puesto que el signo de $f''(x)$ cambia en $x = 0$, la curva seno posee en el origen una tangente de inflexión inclinada a un ángulo de 45 grados con respecto al eje x.

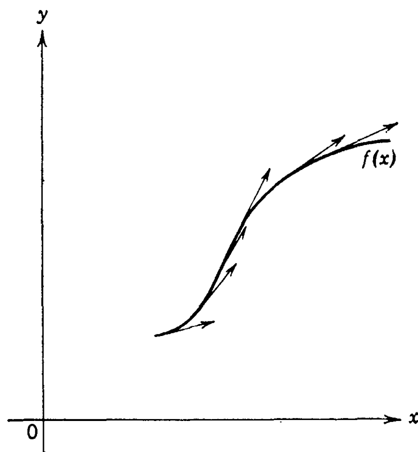


Figura 3.15 Punto de inflexión.

Debe notarse, sin embargo, que pueden existir puntos donde $f''(x) = 0$ y el signo de $f''(x)$ no varíe al crecer x, en tanto que la tangente no corta a la curva sino que permanece por completo a un lado de ésta. Por ejemplo, la curva $y = x^4$ está situada enteramente por encima del eje x, aun cuando la segunda derivada $f''(x) = 12x^2$ se anula para $x = 0$.

b. Máximos y mínimos—Extremos relativos. Puntos estacionarios

Una función $f(x)$ posee un *máximo* en el punto ξ si el valor de f en el punto ξ no es excedido por el valor de f en cualquier otro punto x del dominio de f; esto es, $f(\xi) \geq f(x)$ para todo x donde f esté definida.[1] Análogamente, f posee un *mínimo* en ξ si $f(\xi) \leq f(x)$ para todo x en el dominio. La palabra *extremos* es utilizada para designar tanto a máximos como a mínimos.

---

[1] Se habla de un *punto máximo estricto* ξ si $f(\xi) > f(x)$ para *todo x en el dominio* de f que sea diferente de ξ.

La función $f(x) = \sqrt{1 - x^2}$, por ejemplo, que está definida para $-1 \leq x \leq 1$, posee mínimos en $x = \pm 1$ y un máximo en $x = 0$. Es fácil dar ejemplos de funciones continuas que no poseen máximos o no poseen mínimos. Así, la función $f(x) = 1/(1 + x^2)$ (Fig. 3.8, p. 238) en el dominio $-\infty < x < +\infty$ no posee mínimos; la función $f(x) = 1/x$ definida para $0 < x < \infty$ no posee ningún punto extremo. Sin embargo, recuérdese del capítulo 1, p. 124, el teorema de Weierstrass, de acuerdo con el cual una función continua definida en un intervalo finito cerrado posee siempre un máximo (y análogamente un mínimo).

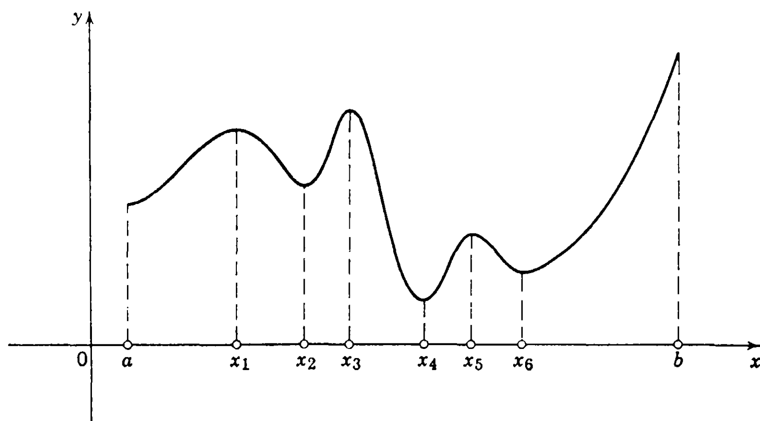


Figura 3.16 Gráfica de una función definida en el intervalo $[a, b]$ con mínimos relativos en $x = a, x_2, x_4, x_6$, máximos relativos en x_1, x_3, x_5, b, máximo absoluto en b, y mínimo absoluto en x_4.

El objetivo aquí es encontrar un medio de localizar los extremos de una función o curva. Este problema, que se encuentra muy frecuentemente en geometría, mecánica, física y otros campos, fue uno de los principales estímulos para el desarrollo del cálculo en el siglo XVII.

El cálculo no proporciona un método directo para determinar los extremos de una función $f(x)$, sino que permite localizar los así llamados puntos *extremos relativos,* entre los cuales los máximos y mínimos reales deben ocurrir. El punto ξ es un máximo *relativo* (mínimo de f si f posee su valor máximo (mínimo) en ξ cuando es comparado no con todos los valores posibles de $f(x)$ sino con sólo aquellos valores de $f(x)$ para x en alguna *vecindad* de ξ. Por una vecindad o entorno del punto ξ se entiende aquí cualquier intervalo abierto $\alpha < x < \beta$ que contiene al punto ξ pero que puede ser arbitrariamente pequeño. Un punto extremo relativo ξ de f es entonces un punto que es un extremo cuando f está restringida

a todos aquellos puntos de su dominio situados suficientemente cerca de ξ.[1] Obviamente, los extremos de una función están incluidos entre los extremos relativos. Para evitar confusión se emplearán los términos *máximos absolutos (mínimos)* para los máximos y mínimos de f en todo su dominio (véase Fig. 3.16).

Geométricamente hablando, los máximos y mínimos relativos, si no están localizados en los puntos extremos del intervalo de definición, son respectivamente las crestas y depresiones de la curva. Una ojeada a la Fig. 3.16 muestra que el valor del máximo relativo en un punto x_5 puede muy bien ser menor que el valor de un mínimo relativo en otro punto x_2. El diagrama sugiere también el hecho de que máximos y mínimos relativos de funciones continuas se alternan: Entre dos máximos relativos sucesivos está siempre localizado un mínimo relativo.

Sea $f(x)$ una función derivable definida en el intervalo cerrado $a \leq x \leq b$. Se ve de inmediato que en un punto extremo relativo que está localizado en el interior del intervalo la tangente a la curva debe ser horizontal. (La demostración rigurosa se da a continuación.) Por lo tanto, la condición

$$f'(\xi) = 0$$

es necesaria para un extremo relativo en un punto ξ tal que $a < \xi < b$. Si, no obstante, $f(\xi)$ es un extremo relativo y ξ coincide con uno de los puntos extremos del intervalo de definición, la ecuación $f'(\xi) = 0$ no necesariamente se cumple. Sólo puede decirse que si el punto extremo izquierdo es un punto máximo relativo (mínimo), la pendiente $f'(a)$ de la curva no puede ser positiva (negativa), mientras que si el punto extremo derecho b es un máximo relativo (mínimo) entonces $f'(b)$ no puede ser negativa (positiva).

Los puntos en los cuales la tangente a la curva $y = f(x)$ es horizontal, que corresponden a las raíces ξ de la ecuación $f'(\xi) = 0$, se denominan los puntos *críticos* o puntos *estacionarios* de f. Todos los extremos relativos de una función derivable f que son puntos *interiores* del dominio de f son puntos estacionarios. Por lo tanto: *un máximo absoluto o un mínimo absoluto de la función coincide ya sea con un punto crítico de la función o bien con un punto frontera de su dominio.* Para localizar los máximos absolutos (mínimos) de la función debemos solamente comparar los valores de f en los puntos críticos y en los dos puntos frontera o extremos del intervalo, y ver cuáles de estos valores son los mayores (menores). Si f no es derivable en un número finito de puntos, se deberá sólo sumar estos puntos a la lista de localizaciones posibles de un extremo y también verifi-

---

[1] La definición rigurosa de un punto de *máximo relativo*, ξ, establecería que existe un intervalo abierto que contiene a ξ tal que $f(\xi) \geq f(x)$ para todo x de este intervalo para el cual f esté definida.

car los valores de f en estos puntos. Así, la labor principal en la determinación de los extremos de una función se reduce a la de encontrar los ceros de la derivada de la función, que usualmente existen en número finito.

Para tomar un ejemplo sencillo, determinemos los valores máximos y mínimos de la función $f(x) = \frac{1}{10}x^6 - \frac{3}{10}x^2$ en el intervalo $-2 \leq x \leq 2$. Aquí los puntos críticos, las raíces de la ecuación $f'(x) = 6(x^5 - x)/10 = 0$, están localizados en $x = 0, +1, -1$. Calculando los valores de f en estos puntos y también en los puntos extremos del intervalo encontramos

x	-2	-1	0	1	2
$f(x)$	5.2	-0.2	0	-0.2	5.2

Es claro que los puntos $x = \pm 1$ representan mínimos relativos, mientras que los máximos relativos ocurren en $x = 0$ y $x = \pm 2$. El valor máximo de la función, alcanzado en los puntos extremos del intervalo, es 5.2; el valor mínimo, alcanzado en los puntos $x = \pm 1$, es -0.2 (véase Fig. 3.17).

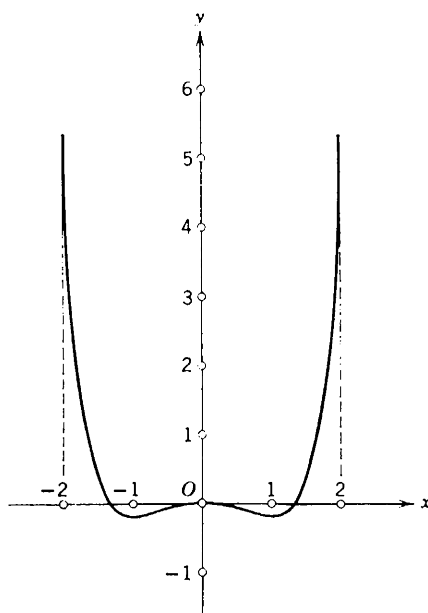


Figura 3.17 $y = (x^6 - 3x^2)/10$.

Sin apelar a la intuición, se puede probar fácilmente mediante métodos puramente analíticos que $f'(\xi) = 0$ siempre y cuando ξ es un punto extremo relativo en el interior del dominio de f, suponiendo f derivable en ξ. (Comparar las consideraciones exactamente análogas para el teorema de Rolle,

p. 196.) Si la función $f(x)$ posee un máximo relativo en el punto ξ, entonces para todos los valores suficientemente pequeños de h diferentes de cero la expresión $f(\xi + h) - f(\xi)$ debe ser negativa o cero. Por lo tanto,

$$\frac{[f(\xi + h) - f(\xi)]}{h} \leq 0$$

para $h > 0$, mientras que

$$\frac{[f(\xi + h) - f(x)]}{h} \geq 0$$

para $h < 0$. Así, si h tiende a cero a través de valores positivos el límite no puede ser positivo, mientras que si h tiende a cero a través de valores negativos el límite no puede ser negativo. Sin embargo, ya que se ha supuesto que la derivada en ξ existe, estos dos límites deben ser iguales entre sí y, precisamente, iguales al valor $f'(\xi)$ que, por consiguiente, solamente puede ser cero; o sea, debe tenerse necesariamente $f'(\xi) = 0$. Una demostración análoga se cumple para un mínimo relativo. La demostración prueba también que si el punto extremo izquierdo $\xi = a$ es un punto máximo relativo (mínimo), entonces por lo menos $f'(a) \leq 0\,[f'(a) \geq 0]$; si el punto extremo derecho b es un punto máximo relativo (mínimo) entonces $f'(b) \geq 0\,[f'(b) \leq 0]$.

La condición $f'(\xi) = 0$ que caracteriza a los puntos críticos no es de ninguna manera suficiente para que existan extremos relativos. Pueden existir puntos en los cuales la derivada se anula, esto es, en los cuales la tangente es horizontal, aun cuando la curva no posee máximo relativo ni mínimo relativo ahí. Esto ocurre si en el punto dado la curva posee una *tangente de inflexión horizontal* que la corta, como en el ejemplo de la función $y = x^3$ en el punto $x = 0$.

El siguiente teorema da las condiciones necesarias y suficientes bajo las cuales un punto crítico es un punto de máximo o mínimo relativo (y, por tanto, un *criterio* para reconocerlo). Se aplica a una función continua, f, que posee una derivada continua f' que se anula en a lo más un número finito de puntos, o, más generalmente, a funciones derivables f para las cuales f' cambia de signo en a lo más un número finito de puntos:

La función $f(x)$ posee un extremo relativo en un punto interior ξ de su dominio si y sólo si la derivada $f'(x)$ cambia de signo cuando x pasa a través de este punto; en particular, la función posee un mínimo relativo si cerca de ξ la derivada es negativa a la izquierda de ξ y positiva a la derecha, mientras que en el caso contrario posee un máximo.

Se demuestra esto rigurosamente utilizando el teorema del valor medio. Primero se observa que a la izquierda y a la derecha de ξ existen intervalos, $\xi_1 < x < \xi$ y $\xi < x < \xi_2$, en cada uno de los cuales $f'(x)$ posee un solo signo, puesto que f' se anula solamente en un número finito de puntos.

(Aquí ξ_1 y ξ_2 pueden ser tomados como los puntos más cercanos a ξ en los cuales f' se anula, si tales puntos existen.) Si los signos de $f'(x)$ en estos dos intervalos son diferentes, entonces $f(\xi + h) - f(\xi) = hf'(\xi + \theta h)$ posee el mismo signo para todos los valores suficientemente pequeños de h, ya sea que h se tome valor positivo o negativo; de modo que ξ es un extremo relativo. Si $f'(x)$ posee el mismo signo en ambos intervalos, entonces $hf'(\xi + \theta h)$ cambia de signo cuando h lo hace, de modo que $f(\xi + h)$ es mayor que $f(\xi)$ en un lado y menor que $f(\xi)$ en el otro lado, y así no existe valor extremo. Nuestro teorema está así probado.

Al mismo tiempo se ve que *el valor $f(\xi)$ es el valor máximo o mínimo de la función en todo intervalo que contiene al punto ξ, en el cual f es derivable y en el cual el único cambio de signo de $f'(x)$ ocurre en ξ mismo.*

El teorema del valor medio, en el que esta demostración se basa, puede seguirse utilizando aún en el caso en que $f(x)$ no es derivable en un punto extremo del intervalo en el cual se aplica, siempre que sea derivable en todos los otros puntos del intervalo; por lo tanto, esta demostración sigue siendo válida si $f'(x)$ no existe en $x = \xi$. Por ejemplo, la función $y = |x|$ posee un mínimo en $x = 0$, puesto que $y' > 0$ para $x > 0$ e $y' < 0$ para $x < 0$ (ver Fig. 2.24, p. 188). La función $y = \sqrt[3]{x^2}$ posee del mismo modo un mínimo en el punto $x = 0$, aun cuando su derivada $\frac{2}{3}x^{-1/3}$ es infinita ahí (ver Fig. 2.27, p. 190).

El método más sencillo para decidir si un punto crítico ξ es un máximo o mínimo relativo involucra la *segunda derivada* en este punto. Es intuitivamente claro que si $f'(\xi) = 0$ entonces f tiene un máximo relativo en ξ si $f''(\xi) < 0$, y tiene un mínimo relativo si $f''(\xi) > 0$. Pues en el primer caso la curva en la vecindad de este punto está situada completamente por debajo de la tangente y en el segundo caso completamente por encima de la tangente. Este resultado se obtiene analíticamente a partir del criterio (teorema) anterior siempre que $f(x)$ y $f'(x)$ sean continuas y que $f''(\xi)$ exista. En efecto, si $f'(\xi) = 0$ y, digamos $f''(\xi) > 0$, se tiene

$$f''(\xi) = \lim_{h\to 0} \frac{f'(\xi + h) - f'(\xi)}{h} = \lim_{h\to 0} \frac{f'(\xi + h)}{h} > 0.$$

Se sigue que $f'(\xi + h)/h > 0$ para todos los $h \neq 0$ que son suficientemente pequeños en valor absoluto; por lo tanto, $f'(\xi + h)$ y h poseen el mismo signo en una vecindad de ξ. Si x está cerca de ξ, la derivada $f'(x)$ debe ser negativa para x a la izquierda de ξ y positiva para x a la derecha de ξ; y esto implica que existe un mínimo relativo en ξ.

La situación es particularmente simple en el caso en que $f''(x)$ es de uno y el mismo signo en todo el intervalo $[a, b]$ en el cual f está definida:

Un punto ξ en el cual f' se anula es un punto de máximo de f si $f''(x) < 0$ en todo el intervalo (o bien si su curva es cóncava), y es un

punto de mínimo de f si en todo el intervalo $f''(x) > 0$ (esto es, si la curva es convexa).[1]

En efecto, si $f''(x) < 0$ la función $f'(x)$ es monótona decreciente, y, por lo tanto, ξ es su único cero. Además, $f' > 0$ para $a \leq x < \xi$, mientras que $f' < 0$ para $\xi < x \leq b$. Por el teorema del valor medio esto implica nuevamente que $f(x) < f(\xi)$ para $x \neq \xi$, de modo que ξ resulta ser un estricto punto de máximo. El mínimo de f debe coincidir con uno de los puntos frontera, a, b, puesto que no existe otro punto crítico aparte de ξ. El mismo argumento se aplica cuando $f'' > 0$ en el intervalo.

Ejemplos

Ejemplo 1. De entre todos los triángulos con base y área dadas determinar aquél con mínimo perímetro.

Para resolver este problema se toma al eje x a lo largo de la base dada AB y el punto medio de AB como el origen (Fig. 3.18). Si C es el vértice del triángulo, h su altura (la cual está determinada por el área y la base), y (x, h) son las coordenadas del vértice, entonces la suma de los dos lados AC y BC del triángulo está dada por

$$f(x) = \sqrt{(x+a)^2 + h^2} + \sqrt{(x-a)^2 + h^2},$$

donde $2a$ es la longitud de la base. De aquí se obtiene

$$f'(x) = \frac{x+a}{\sqrt{(x+a)^2 + h^2}} + \frac{x-a}{\sqrt{(x-a)^2 + h^2}},$$

$$f''(x) = \frac{-(x+a)^2}{\sqrt{[(x+a)^2 + h^2]^3}} + \frac{1}{\sqrt{(x+a)^2 + h^2}}$$

$$+ \frac{-(x-a)^2}{\sqrt{[(x-a)^2 + h^2]^3}} + \frac{1}{\sqrt{(x-a)^2 + h^2}}$$

$$= \frac{h^2}{\sqrt{[(x+a)^2 + h^2]^3}} + \frac{h^2}{\sqrt{[(x+a)^2 + h^2]^3}}.$$

Vemos de inmediato que: (1) $f'(0)$ se anula, y (2) $f''(x)$ es siempre positiva; por lo tanto, en $x = 0$ existe un valor mínimo (ver pp. 264 y 265). Este valor mínimo está dado según esto por el triángulo isósceles.

Análogamente se encuentra que de entre todos los triángulos con perímetro y base dados el triángulo isósceles posee el área máxima.

Ejemplo 2. Encontrar un punto sobre una recta dada tal que la suma de sus distancias a dos puntos fijos sea un mínimo.

---

[1] *También se dice, brevemente, que el punto es un "máximo (mínimo) de $f(x)$" (N. del R.).*

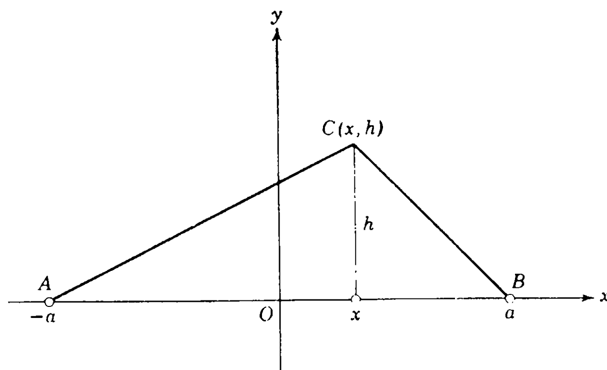


Figura 3.18.

Supónganse dados la recta y dos puntos fijos cualesquiera A y B en el mismo lado de la recta. Se desea encontrar un punto P sobre la línea recta, tal que la distancia $PA + PB$ tenga el mínimo valor posible.[1]

Se toma la recta dada como el eje x y se utiliza la notación de la Fig. 3.19. Entonces, la distancia en cuestión está dada por

$$f(x) = \sqrt{x^2 + h^2} + \sqrt{(x-a)^2 + h_1^2},$$

y se obtiene

$$f'(x) = \frac{x}{\sqrt{x^2 + h^2}} + \frac{x-a}{\sqrt{(x-a)^2 + h_1^2}},$$

$$f''(x) = \frac{h^2}{\sqrt{(x^2 + h^2)^3}} + \frac{h_1^2}{\sqrt{[(x-a)^2 + h_1^2]^3}}.$$

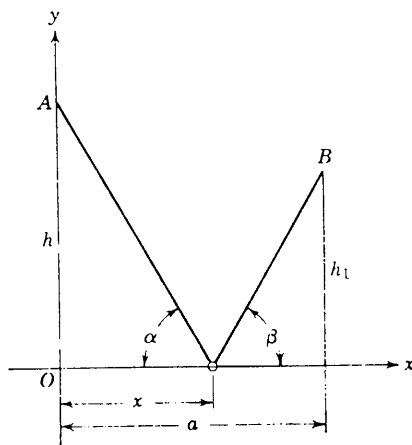


Figura 3.19 Ley de reflexión.

[1] Si A y B están situados en lados opuestos de la recta, obviamente P es, precisamente, la intersección de la recta con el segmento AB.

La ecuación $f'(\xi) = 0$ significa

$$\frac{\xi}{\sqrt{\xi^2 + h^2}} = \frac{a - \xi}{\sqrt{(\xi - a)^2 + h_1^2}},$$

o bien

$$\cos \alpha = \cos \beta;$$

y, por lo tanto, las dos rectas PA y PB deben formar ángulos iguales con la recta dada. El signo positivo de $f''(x)$ muestra que realmente tenemos un valor mínimo.

La solución de este problema está conectada muy de cerca con la ley de reflexión óptica. Por un principio importante de la óptica, conocido

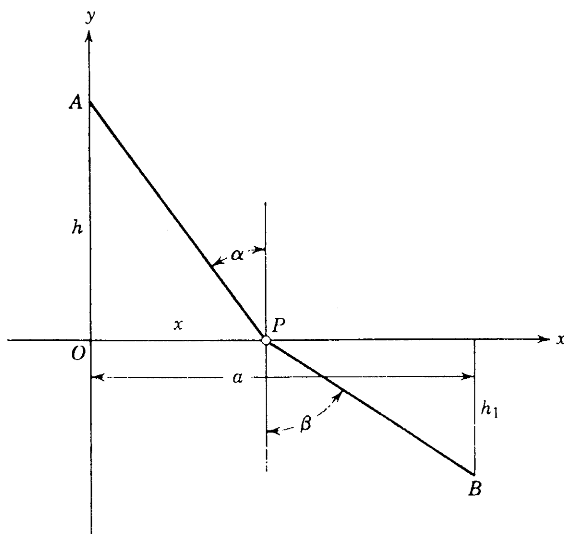


Figura 3.20 Ley de refracción.

como *"principio del tiempo mínimo"* de Fermat, la trayectoria de un rayo de luz está determinada por la propiedad de que el tiempo que tarda la luz en ir de un punto A a un punto B bajo las condiciones dadas debe ser el menor posible. Si se impone la condición de que un rayo de luz en su recorrido de A a B pase por algún punto de una recta dada (digamos, sobre un espejo), se ve que el tiempo mínimo se logrará a lo largo del rayo para el cual el "ángulo de incidencia" sea igual al "ángulo de reflexión".

Ejemplo 3. La Ley de Refracción.[1] Considérense dados dos puntos A y B en lados opuestos del eje x. ¿Cuál es la trayectoria de A a B que

[1] En tanto que los ejemplos anteriores pueden ser tratados también mediante la geometría elemental, el presente no puede ser fácilmente resuelto sin el cálculo.

requiere el menor tiempo posible si la velocidad en un lado del eje x es c_1 y en el otro lado es c_2?

Claramente, esta "trayectoria mínima" debe consistir de dos porciones de líneas rectas que se encuentran una a la otra en un punto P sobre el eje x. Utilizando la notación de la Fig. 3.20 se obtienen las dos expresiones $\sqrt{h^2 + x^2}$ y $\sqrt{h_1^2 + (a - x)^2}$ para las longitudes PA, PB, respectivamente; y el tiempo, T, de viaje a lo largo de esta trayectoria se encuentra dividiendo las longitudes de los dos segmentos por las correspondientes velocidades y después sumando:

$$T = f(x) = \frac{1}{c_1}\sqrt{h^2 + x^2} + \frac{1}{c_2}\sqrt{h_1^2 + (a - x)^2}.$$

Diferenciando, se obtiene

$$f'(x) = \frac{1}{c_1}\frac{x}{\sqrt{h^2 + x^2}} - \frac{1}{c_2}\frac{a - x}{\sqrt{h_1^2 + (a - x)^2}},$$

$$f''(x) = \frac{1}{c_1}\frac{h^2}{\sqrt{(h^2 + x^2)^3}} + \frac{1}{c_2}\frac{h_1^2}{\sqrt{[h_1^2 + (a - x)^2]^3}}.$$

Como fácilmente se ve por la Fig. 3.20, la ecuación $f'(x) = 0$, esto es, la ecuación

$$\frac{1}{c_1}\frac{x}{\sqrt{h^2 + x^2}} = \frac{1}{c_2}\frac{a - x}{\sqrt{h_1^2 + (a - x)^2}},$$

es equivalente a la condición $(1/c_1)$ sen $\alpha = (1/c_2)$ sen β, o bien,

$$\frac{\operatorname{sen} \alpha}{\operatorname{sen} \beta} = \frac{c_1}{c_2}.$$

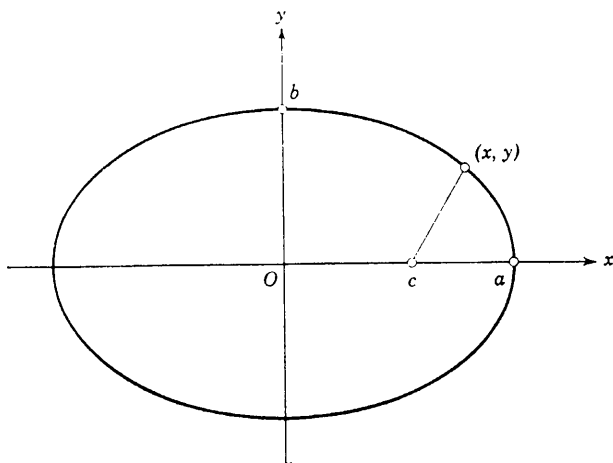


Figura 3.21 Punto de una elipse que está a la mínima distancia de otro sobre el eje mayor.

El lector debe verificar el hecho de que existe sólo *un* punto que satisface la condición, y que este punto proporciona en realidad el valor mínimo requerido.

El significado físico de nuestro ejemplo está precisamente dado por el principio óptico del tiempo mínimo. Un rayo de luz que viaja entre dos puntos describe una trayectoria de tiempo mínimo. Si c_1 y c_2 son las velocidades de la luz en dos medios ópticos distintos, la trayectoria luminosa estará dada por nuestro resultado, el cual es una forma de la *Ley de Snell de la refracción.*

Ejemplo 4. Encontrar el punto de una elipse que está más cerca que cualquier otro de un punto dado sobre su eje mayor (Fig. 3.21).

Tomando la ecuación de la elipse en la forma

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \qquad (b < a),$$

y el punto dado sobre el eje mayor como $(c, 0)$, se encuentra para la distancia de cualquier punto (x, y) sobre la elipse al punto $(c, 0)$ la expresión

$$d = \sqrt{(x - c)^2 + b^2(1 - x^2/a^2)},$$

donde $-a \leq x \leq a$. La función $f(x) = d^2$ es convexa $(f'' > 0)$. Posee un mínimo para el mismo x que d. El único punto crítico de f está en $x = c/(1 - b^2/a^2)$. Si este punto está situado en el dominio de d, representa un punto de mínimo; si no, el mínimo de d corresponde al punto extremo del eje mayor que esté más cercano a c. De acuerdo con esto, se encuentran para la mínima distancia los valores

$$d = b\sqrt{1 - \frac{c_2}{a^2 - b^2}} \qquad \text{si} \quad |c| \leq a\left(1 - \frac{b^2}{a^2}\right),$$

$$d = a - |c| \qquad \text{si} \quad |c| \geq a\left(1 - \frac{b^2}{a^2}\right).$$

*3.7 El orden de magnitud de las funciones

Las diferencias entre los comportamientos de las funciones para valores grandes del argumento conducen a la noción de *orden de magnitud.* Debido a su gran importancia este asunto merece una breve discusión aquí, aunque no esté directamente conectado con la idea de la integral o de la derivada.

a. El concepto de orden de magnitud. Los casos más simples

Si la variable x crece más allá de toda cota, entonces, para $\alpha > 0$, las funciones x^α, $\log x$, e^x, $e^{\alpha x}$ crecen también más allá de toda cota. Estas, sin

embargo, se incrementan en formas esencialmente diferentes. Por ejemplo, la función x^3 se torna "infinita en un orden superior" [1] al de x^2; y por esto se entiende que: conforme x crece, el cociente x^3/x^2 se incrementa más allá de toda cota. Análogamente, la función x^α se torna infinita en un orden superior al de x^β si $\alpha > \beta > 0$, etc.

En forma general se dice de dos funciones, $f(x)$ y $g(x)$, cuyos valores absolutos se incrementan con x más allá de toda cota, que $f(x)$ se ***torna infinita de un orden superior al de*** $g(x)$ si para $x \to \infty$ el cociente $|f(x)/g(x)|$ se incrementa más allá de toda cota; y diremos que $f(x)$ se torna infinita en un orden menor que el de $g(x)$ si el cociente $|f(x)/g(x)|$ tiende a cero conforme x crece; y se dirá que las dos funciones se tornan *infinitas del mismo orden de magnitud* si, al crecer x, el cociente $|f(x)/g(x)|$ posee un límite diferente de cero o, al menos, permanece entre dos cotas fijas positivas. Por ejemplo, la función $ax^3 + bx^2 + c = f(x)$, donde $a \neq 0$, será del mismo orden de magnitud que la función $x^3 = g(x)$, pues el cociente $|f(x)/g(x)| = |(ax^3 + bx^2 + c)/x^3|$ posee el límite $|a|$ conforme $x \to \infty$. Por otra parte, la función $x^3 + x + 1$ se torna infinita de un orden de magnitud superior al de la función $x^2 + x + 1$.

Una suma de dos funciones $f(x)$ y $\phi(x)$, donde $f(x)$ es de orden de magnitud superior al de $\phi(x)$, posee el mismo orden de magnitud que $f(x)$.

Pues $|(f(x) + \phi(x))/f(x)| = |1 + \phi(x)/f(x)|$, y, por hipótesis, esta expresión tiende a uno cuando x crece más allá de toda cota.

b. Los órdenes de magnitud de la función exponencial y del logaritmo

Se podría estar tentado a medir el orden de magnitud de las funciones mediante una escala, asignando a la cantidad x el orden de magnitud uno y a la potencia x^α $(\alpha > 0)$ el orden de magnitud α. Un polinomio de grado n tendría entonces obviamente el orden de magnitud n; una función racional, el grado de cuyo numerador es superior en h al del denominador, tendría el orden de magnitud h.

Salta a la vista, sin embargo, que cualquier intento para describir el orden de magnitud de funciones arbitrarias mediante tal escala debe fallar. Esto es así porque existen funciones que se tornan infinitas en un orden superior al de la potencia x^α de x, no importa cuán grande se elija α; además, existen funciones que se tornan infinitas en un orden menor que el de la potencia x^α, no importa cuán pequeño se elija el número positivo α. Estas funciones, por lo tanto, no encajan en nuestra escala.

Sin entrar en una teoría detallada, se establece el siguiente teorema.

---

[1] O "infinita de orden superior" al de x^2.

TEOREMA. *Si a es un número arbitrario mayor que uno, entonces el cociente a^x/x tiende a infinito cuando x crece.*

DEMOSTRACIÓN. Para probar esto se construye la función

$$\phi(x) = \log \frac{a^x}{x} = x \log a - \log x;$$

y es obvio que será suficiente mostrar que $\phi(x)$ crece más allá de toda cota si x tiende a $+\infty$. Con este objeto, se considera la derivada

$$\phi'(x) = \log a - \frac{1}{x}$$

y se observa que para $x \geq c = 2/\log a$ no es menor que el número positivo $\frac{1}{2}\log a$. Por lo tanto, se sigue que para $x \geq c$ es

$$\phi(x) - \phi(c) = \int_c^x \phi'(t)\,dt \geq \int_c^x \tfrac{1}{2}\log a\,dt \geq \tfrac{1}{2}(x - c)\log a,$$

$$\phi(x) \geq \phi(c) + \tfrac{1}{2}(x - c)\log a,$$

y el miembro derecho se torna infinito para $x \to \infty$.

Se da en seguida una segunda demostración de este importante teorema: con $\sqrt{a} = b = 1 + h$, se tiene $b > 1$ y $h > 0$. Sea n el entero tal que $n \leq x \leq n + 1$; y puede tomarse $x > 1$, de modo que $n \geq 1$. Aplicando el lema de la p. 87 se tiene

$$\sqrt{\frac{a^x}{x}} = \frac{b^x}{\sqrt{x}} = \frac{(1+h)^x}{\sqrt{x}} > \frac{(1+h)^n}{\sqrt{n+1}} > \frac{1+nh}{\sqrt{n+1}} > \frac{nh}{\sqrt{2n}} = \frac{h}{\sqrt{2}}\sqrt{n},$$

de modo que

$$\frac{a^x}{x} > \frac{h^2}{2}\cdot n,$$

y, por lo tanto, el cociente tiende a infinito con x.

De este hecho que se acaba de probar se siguen muchos otros. Por ejemplo: para todo índice positivo α y todo número $a > 1$, el cociente a^x/x^α tiende a infinito cuando x crece. Esto es, vale el siguiente

TEOREMA. *La función exponencial se torna infinita de un orden de magnitud superior al de cualquier potencia de x.*[1]

Para demostrarlo sólo se necesita mostrar que la α-ésima raíz de la expresión, esto es,

$$\frac{a^{x/\alpha}}{\alpha} = \frac{1}{\alpha}\frac{a^{x/\alpha}}{x/\alpha} = \frac{1}{\alpha}\cdot\frac{a^y}{y} \qquad \left(y = \frac{x}{\alpha}\right),$$

[1] También se dice frecuentemente; por brevedad: "de orden superior a cualquier potencia de x". Otra expresión equivalente es: "a^x crece a infinito *más rápidamente* que x^α cuando $x \to \infty$".

tiende a infinito. Esto, sin embargo, se sigue inmediatamente a partir del teorema anterior cuando x es reemplazado por $y = x/\alpha$.

De un modo análogo se demuestra el siguiente teorema. Para todo valor positivo de α el cociente $(\log x)/x^{\alpha}$ tiende a cero cuando $x \to \infty$; esto es

TEOREMA. *El logaritmo se torna infinito de un orden menor de magnitud que el de cualquier potencia positiva arbitrariamente pequeña de x.*[1]

DEMOSTRACIÓN. La demostración se obtiene de inmediato si se expresa $\log x = y$ y así el cociente es transformado en $y/e^{\alpha y}$. Se hace entonces $e^{\alpha} = a$; entonces $a > 1$, y el cociente y/a^{y} se aproxima a cero cuando y tiende a infinito. Puesto que y se aproxima a infinito conforme x lo hace, el teorema está demostrado.[2]

Sobre la base de estos resultados pueden construirse funciones de un orden de magnitud muy superior al de la función exponencial, y otras funciones de un orden de magnitud bastante más pequeño que el del logaritmo. Por ejemplo, la función $e^{(e^x)}$ es de un orden superior al de la función exponencial, y la función $\log \log x$ es de un orden menor que el del logaritmo; además puede iterarse este proceso tantas veces como se quiera, reiterando los símbolos u operaciones $e^{(\ )}$ o log tantas veces como se desee.

Todas las funciones x, $\log x$, $\log(\log x)$, $\log[\log(\log x)]$, etc., eventualmente se tornan arbitrariamente grandes para x suficientemente grande, pero con creciente lentitud. Tomando, por ejemplo, para x el enorme número $x = 10^{100}$ se encuentra que $\log x$ es más o menos 230, mientras que $\log(\log x)$ es sólo 5.4, aproximadamente.

c. Comentarios generales

Estas consideraciones muestran que no es posible asignar a todas las funciones números definidos, como órdenes de magnitud, de modo que de dos funciones, aquélla con el orden de magnitud más alto tenga un número superior. Si, por ejemplo, la función x es de orden de magnitud uno y la función $x^{1+\varepsilon}$ de orden de magnitud $1 + \varepsilon$, entonces la función $x \log x$ debe ser de un orden de magnitud mayor que uno y menor que $1 + \varepsilon$, no importa cuán pequeño sea elegido ε. Pero no existe tal número.

Además, es fácil ver que las funciones no tienen que poseer un orden relativo de magnitud claramente definido. Por ejemplo, la función

---

[1] También: "crece *menos rápidamente que* x^{α}, cuando $x \to \infty$, por pequeño que sea α". (Cf. Nota al pie de la página anterior.) *(N. del R.)*

[2] Puede sugerirse otra demostración sencilla: Para $x > 1$ y $\varepsilon > 0$

$$\log x = \int_1^x \frac{d\xi}{\xi} < \int_1^x \xi^{\varepsilon-1}\, d\xi = \frac{1}{\varepsilon}(x^{\varepsilon} - 1) \cdot$$

y, si se elige ε igual a α y se dividen ambos miembros de esta desigualdad por x^{α}, entonces se sigue que $(\log x)/x^{\alpha} \to 0$ cuando $x \to \infty$.

$[x^2(\operatorname{sen} x)^2 + x + 1]/[x^2(\cos x)^2 + x]$ no se aproxima a un límite definido cuando x crece; por el contrario, para $x = n\pi$ (donde n es un entero) el valor es $1/n\pi$, mientras que para $x = (n + \frac{1}{2})\pi$ es $(n + \frac{1}{2})\pi + 1 + 1/(n + \frac{1}{2})\pi$. Aunque tanto el numerador como el denominador se tornan infinitos, el cociente no permanece entre cotas positivas, ni tiende a cero ni tiende a infinito. El numerador, por lo tanto, no es del mismo orden que el denominador, ni de orden mayor o menor. Esta situación aparentemente alarmante significa solamente que las definiciones anteriores no están diseñadas de tal forma que se puedan comparar dos funciones cualesquiera. Este no es un defecto. No se desea aquí comparar los órdenes de tales funciones como el numerador y el denominador mencionados; pues el conocimiento del valor de uno de ellos no da ninguna información útil acerca del otro.

d. El orden de magnitud de una función en la vecindad de un punto arbitrario

Así como se pueden comparar los comportamientos de funciones para $x \to \infty$, pueden compararse también funciones que se tornan infinitas en el punto finito $x = \xi$.

Se dice que la función $f(x) = 1/|x - \xi|$ se torna *infinita del primer orden* en el punto $x = \xi$, y, correspondientemente, que la función $1/|x - \xi|^\alpha$ se torna *infinita de orden* α, siempre que α sea positiva.

Se reconoce entonces que la función $e^{1/|x-\xi|}$ se torna infinita de orden superior para $x \to \xi$, y la función $\log |x - \xi|$ infinita de menor orden que el de todas estas potencias; esto es, que las relaciones límite

$$\lim_{x \to \xi} (|x - \xi|^\alpha \cdot e^{1/|x-\xi|}) = \infty \qquad \text{y} \qquad \lim_{x \to \xi} (|x - \xi|^\alpha \cdot \log |x - \xi|) = 0$$

se cumplen.

Para confirmar esto, solamente se hace $1/|x - \xi| = y$; y nuestros asertos se reducen entonces al conocido teorema de la p. 271, puesto que

$$|x - \xi|^\alpha \cdot e^{1/|x-\xi|} = \frac{e^y}{y^\alpha} \qquad \text{y} \qquad |x - \xi|^\alpha \cdot \log |x - \xi| = -\frac{\log y}{y^\alpha},$$

y que y crece más allá de toda cota conforme x tiende a ξ. (El método de reducir el comportamiento en un punto ξ al comportamiento en el infinito mediante la substitución $1/|x - \xi| = y$ resulta ser frecuentemente útil.)

e. El orden de magnitud (pequeñez) de una función que tiende a cero

Tal como se buscó describir la aproximación de una función al infinito por medio del concepto de orden de magnitud, puede especificarse también

la forma en la cual una función se aproxima a cero. Se dice que cuando $x \to \infty$ la cantidad $1/x$ *se anula de (con) primer orden;* y que la cantidad $x^{-\alpha}$, donde α es positiva, se *anula de orden* α.[1] Se encuentra nuevamente que la *función* $1/\log x$ *se anula con orden menor que el de una potencia arbitraria* $x^{-\alpha}$; esto es, para todo α positivo la relación

$$\lim_{x \to \infty} (x^{-\alpha} \cdot \log x) = 0$$

se cumple.

De la misma manera se dice que para $x = \xi$ la cantidad $x - \xi$ se anula en primer orden y la cantidad $|x - \xi|^{\alpha}$ en orden α. Con nuestros resultados es fácil probar las relaciones

$$\lim_{x \to 0} (|x|^{\alpha} \cdot \log |x|) = 0, \qquad \lim_{x \to 0} (|x|^{-\alpha} \cdot e^{-1/|x|}) = 0,$$

las cuales son expresadas usualmente como sigue:

La función $1/\log |x|$ *se anula cuando* $x \to 0$ *con un orden menor que el de cualquier potencia positiva de* x*; y la función exponencial* $e^{-1/|x|}$ *se anula con un orden superior al de cualquier potencia positiva de* x.[2]

f. Las notaciones "O" y "o" para órdenes de magnitud

Una forma conveniente de indicar que una función $f(x)$ es de menor orden de magnitud que una función $g(x)$ es la de escribir $f = o(g)$. Esta ecuación simbólica significa solamente que el cociente f/g posee el límite cero, y puede ser utilizada con igual ventaja para funciones que se anulan o se tornan infinitas y para argumentos x que tienden a infinito o se aproximan a un valor ξ.[3]

Muchos de los resultados de la seccción anterior pueden reescribirse en esta notación; por ejemplo,

$x^{\alpha} = o(x^{\beta})$	para $\alpha < \beta$	conforme $x \to \infty$
$\log x = o(x^{\alpha})$	para $\alpha > 0$	conforme $x \to \infty$
$e^{-x} = o(x^{-\alpha})$	conforme $x \to \infty$	
$e^{-1/x} = o(x^{\alpha})$	conforme $x \to 0$	con valores positivos
$\log \lvert x \rvert = o(1/x)$	conforme $x \to 0$	
$1 - \cos x = o(x)$	conforme $x \to 0$.	

---

[1] También se dice que "$1/x$ *es infinitésimo de primer orden* y $x^{-\alpha}$ $(\alpha > 0)$ *es infinitésimo de orden* α". *(N. del R.)*

[2] O sea: "La función $1/\log |x|$, cuando $x \to 0$, se torna un *infinitésimo de orden inferior* al de cualquier potencia $x^{\alpha} (\alpha > 0)$; y la función $e^{-1/|x|}$ se torna un *infinitésimo de orden superior* al de cualquier potencia $x^{\alpha} (\alpha > 0)$". *(N. del R.)*

[3] La letra o es elegida para sugerir la palabra "orden". Obsérvese que la relación $f = o(g)$ para una g que se anula significa que f se anula con un orden superior.

Esta notación, introducida por E. Landau, es útil para indicar el orden de magnitud del error en una fórmula de aproximación. Por ejemplo,

$$\frac{1}{\sqrt{1+4x^2}} = \frac{1}{2x} + o\left(\frac{1}{x}\right) \qquad \text{para } x \to \infty$$

es la notación para la relación

$$\lim_{x\to\infty} \frac{\dfrac{1}{\sqrt{1+4x^2}} - \dfrac{1}{2x}}{1/x} = 0.$$

Análogamente, la relación entre incremento y diferencial de una función f que posee derivada en un punto x puede ser escrita en la forma

$$f(x+h) - f(x) = hf'(x) + o(h) \qquad \text{para} \quad h \to 0.$$

Igualmente útil es la notación simbólica $f = O(g)$ usada para indicar que $f(x)$ es *a lo más* del orden de magnitud de $g(x)$, esto es, que el cociente $f(x)/g(x)$ *está acotado* para los valores de x en cuestión.[1] El uso del símbolo O es, además, muy flexible. Así, la frase "$f = O(g)$ para $x \to \infty$" significa que el cociente f/g está acotado para todo x suficientemente grande, como en

$$\sqrt{10x - 1} = O(\sqrt{x}) \qquad \text{para} \quad x \to \infty.$$

Análogamente, "$f = O(g)$ para $x \to \xi$" significa que f/g está acotada en una vecindad suficientemente pequeña del punto $x = \xi$, como en

$$e^x - 1 = O(x) \qquad \text{para } x \to 0.$$

Más generalmente, puede utilizarse la ecuación $f = O(g)$ para indicar la acotación de f/g en cualquier dominio del eje x sin que se requiera que x se aproxime a un límite. Así

$$\log x = O(x) \qquad \text{para } x > 1,$$

$$x = O(\operatorname{sen} x) \qquad \text{para } |x| < \frac{\pi}{2}.$$

Algunos de los primeros ejemplos que involucran el símbolo o pueden ser ahora refinados indicándose una mejor estimación del error con la ayuda del símbolo O. Así se tiene para una función f para la cual f'' está definida y es continua:

$$f(x+h) - f(x) = hf'(x) + O(h^2) \qquad \text{para } h \to 0.$$

---

[1] Nótese que $f = O(g)$ *no* significa que f/g tenga el límite uno o que el cociente tenga necesariamente un límite, cualquiera que sea.

Otros ejemplos son

$$\frac{1}{\sqrt{1+4x^2}} = \frac{1}{2x} + O\left(\frac{1}{x^2}\right),$$

$$\cos x = 1 + O(x^2) \qquad \text{para todo } x.$$

Las mismas notaciones pueden utilizarse para *sucesiones* a_n, con el índice n que tienda a infinito. En lo siguiente se encontrarán algunos ejemplos interesantes de tales fórmulas "asintóticas" con un término de error de orden superior (ver la fórmula de Stirling para $n!$ Cap. 6, A.1). Una famosa ley asintótica,[1] ya mencionada en el capítulo 1, p. 80, afirma que el número $\pi(n)$ de primos menores que n está dado aproximadamente por $n/(\log n)$. Aquí el orden de magnitud del error ha sido encontrado también y se tiene, más precisamente, el resultado

$$\pi(n) = \frac{n}{\log n} + O\left(\frac{n}{\log^2 n}\right).$$

Apéndice

La dificultad para apreciar un desarrollo riguroso del cálculo surge de un dilema básico: Aunque los procedimientos y conceptos fundamentales como continuidad, suavidad, etc., son motivados por apremiantes necesidades intuitivas, deben hacerse precisos para que tengan significado lógico y las resultantes definiciones rigurosas puedan cubrir fenómenos que están más allá de aquéllos de carácter intuitivo. Así, el concepto riguroso de continuidad requiere inevitablemente un grado de abstracción que no está totalmente reflejado en la noción intuitiva de una curva conexa; y el concepto de derivabilidad es más restrictivo y más abstracto que la vaga idea que sugiere la *suavidad* de una curva. Discrepancias de este tipo son inevitables y pueden arruinar la paciencia y el entendimiento de un principiante o de alguien para quien la sutileza lógica no es de principal interés. Sin embargo, deseamos hacer más clara para el lector la necesidad de dicha precisión, mostrando que, quizá inesperadamente, la precisión y el refinamiento son requeridos aún para los ejemplos más sencillos e intuitivamente comprensibles.

A.1 Algunas funciones especiales

Como regla, tales ejemplos pueden no estar dados en términos de expresiones analíticas sencillas (véanse Figs. 1.28, p. 62 y 1.30, p. 63). Aquí,

[1] La demostración no puede darse en este libro. Véase: A. E. Ingham, *The Distribution of Primes,* Cambridge University Press, 1932.

sin embargo, se desea representar varios tipos de discontinuidades y fenómenos inesperados o "anormales" mediante expresiones muy sencillas construidas a partir de las funciones elementales. Se comienza con un ejemplo en el cual no se presenta ninguna discontinuidad.

a. La función $y = e^{-1/x^2}$

Esta función (ver Fig. 3.22) está definida en primera instancia solamente para valores de x diferentes de cero, y obviamente posee el límite cero cuando $x \to 0$. En efecto, mediante la transformación $1/x^2 = \xi$ la función se vuelve $y = e^{-\xi}$, y $\lim e^{-\xi} = 0$. Por lo tanto, es natural extender la función de modo que sea continua para $x = 0$ definiendo el valor de la función en el punto $x = 0$ como $y(0) = 0$.

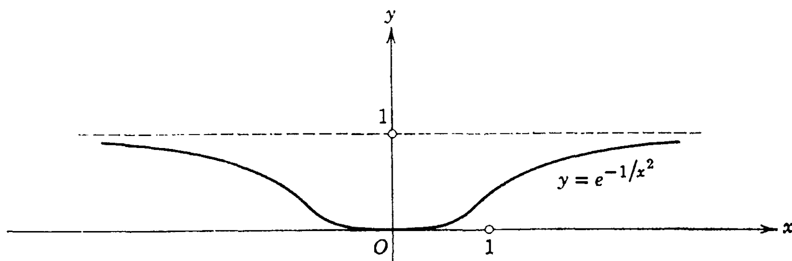


Figura 3.22.

Por la regla de la cadena, la derivada de la función para $x \neq 0$ es $y' = -(2/x^3)e^{-1/x^2} = 2\xi^{3/2}e^{-\xi}$. Si x tiende a cero esta derivada posee también el límite cero, como se deduce inmediatamente de la p. 271. En el punto $x = 0$ la derivada

$$y'(0) = \lim_{h \to 0} \frac{y(h) - y(0)}{h} = \lim_{h \to 0} \frac{e^{-1/h^2}}{h}$$

puede ser también definida en forma continua como cero.

Para las derivadas superiores en $x \neq 0$, obviamente se obtiene siempre el producto de la función e^{-1/x^2} y un polinomio en $1/x$, y el paso al límite $x \to 0$ conduce siempre al límite cero. Por lo tanto, todas las derivadas superiores se anulan, como y', en el punto $x = 0$.

Así, la función es continua en todas partes y derivable tantas veces como se desee; y, aún más, en el punto $x = 0$ se anula con todas sus derivadas, pero no se anula idénticamente. Posteriormente se podrá observar (Apéndice I.1 del capítulo 5) cuán notable o "anormal" es este comportamiento.

b. La función $y = e^{-1/x}$

Como se ve fácilmente, para valores positivos de x esta función se comporta de la misma manera que la función que acabamos de tratar: si x tiende a cero por un lado, con valores positivos, la función tiende a cero, y lo mismo es cierto para todas sus derivadas. Si se define el valor de la función en $x = 0$ como $y(0) = 0$, todas las derivadas laterales derechas en el punto $x = 0$ poseen el valor cero. Otra situación muy diferente se tiene cuando x tiende a cero a través de valores negativos; pues entonces la función y todas sus derivadas se vuelven infinitas, y las derivadas laterales izquierdas no existen en el punto $x = 0$. En el punto $x = 0$, por lo tanto, la función posee una clase notable de discontinuidad bastante diferente de las discontinuidades infinitas de las funciones racionales consideradas en las pp. 60, 61 (ver Fig. 3.23).

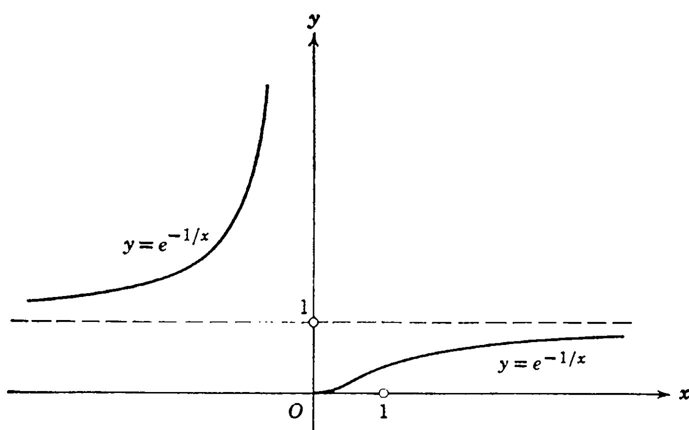


Figura 3.23.

c. La función $y = \tanh \dfrac{1}{x}$

Como se ha visto ya en la p. 88, las funciones con discontinuidades de "salto" pueden obtenerse a partir de funciones elementales mediante el paso al límite. La función exponencial, definida en la p. 172, junto con el principio de la composición de funciones, nos da otro método para construir funciones con tales discontinuidades a partir de funciones elementales, sin ningún otro proceso límite adicional. Un ejemplo de esto es la función

$$y = \tanh \frac{1}{x} = \frac{e^{1/x} - e^{-1/x}}{e^{1/x} + e^{-1/x}}$$

y su comportamiento en el punto $x = 0$. La función en primera instancia no está definida en este punto. Si x se acerca al punto $x = 0$ a través de valores positivos, obviamente se obtiene el límite 1; si, por otra parte, se aproxima al punto $x = 0$ a través de valores negativos, se obtiene el límite -1. Este punto $x = 0$ es así un punto de discontinuidad de salto; conforme x crece pasando por 0 el valor de la función salta dos unidades (ver Fig. 3.24). Por otra parte, la derivada

$$y' = -\frac{1}{\cosh^2(1/x)}\frac{1}{x^2}$$

$$= -\frac{1}{x^2}\frac{4}{(e^{1/x} + e^{-1/x})^2}$$

se aproxima al límite cero desde ambos lados, como surge fácilmente[1] a partir de la sección 3.7, p. 270.

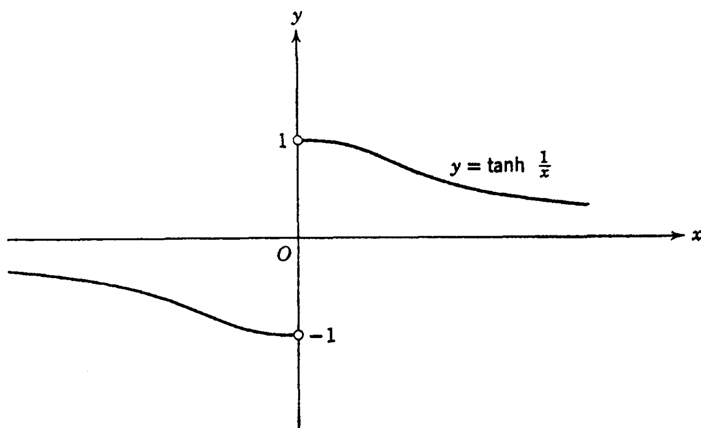


Figura 3.24.

d. La función $y = x \tanh 1/x$

En el caso de la función

$$y = x \tanh\frac{1}{x} = x\,\frac{e^{1/x} - e^{-1/x}}{e^{1/x} + e^{-1/x}}$$

la discontinuidad anterior es eliminada por el factor x. Esta función tiene el límite cero cuando $x \to 0$ desde cualquier lado, de modo que se puede nuevamente definir en forma apropiada $y(0)$ como igual a cero. Esta función es entonces continua en $x = 0$, pero su primera derivada

---

[1] Otro ejemplo donde aparece una discontinuidad "de salto" está dado por la función $y = \arctan 1/x$ cuando $x \to 0$.

$$y' = \tanh\frac{1}{x} - \frac{1}{x}\frac{1}{\cosh^2(1/x)}$$

posee justamente la misma clase de discontinuidad del ejemplo anterior. La gráfica de la función es una curva con una esquina (ver Fig. 3.25); y en el punto $x = 0$ la función no posee una derivada, sino una derivada lateral derecha con el valor $+1$ y una derivada lateral izquierda con el valor -1.

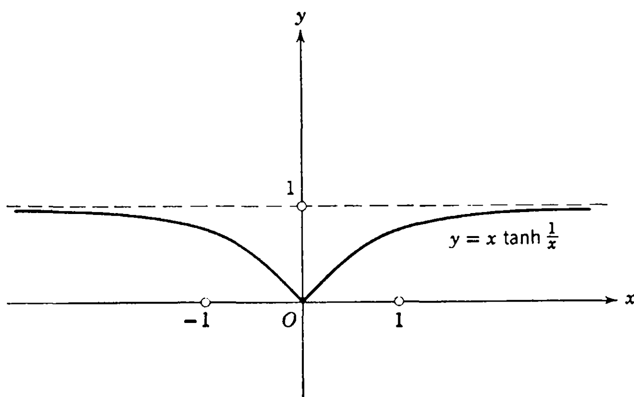


Figura 3.25.

e. La función $y = x \operatorname{sen} 1/x,\ y(0) = 0$

Se ha visto ya que esta función no está compuesta por un número finito de partes monótonas—o, lo que es lo mismo, no es "seccionalmente" monótona o monótona "por pedazos"—pero que, sin embargo, es continua (p. 63 y Fig. 1.31). Su primera derivada

$$y' = \operatorname{sen}\frac{1}{x} - \frac{1}{x}\cos\frac{1}{x} \qquad (x \neq 0),$$

por el contrario, posee una discontinuidad en $x = 0$; pues conforme x tiende a cero esta derivada oscila continuamente entre curvas acotadas, una positiva y una negativa, las cuales a su vez tienden a $+\infty$ y $-\infty$ respectivamente. En el punto $x = 0$ el cociente incremental es $[y(h) - y(0)]/h = \operatorname{sen}(1/h)$; y puesto que este valor oscila hacia arriba y hacia abajo entre 1 y -1 un número infinito de veces conforme $h \to 0$, la función no posee una derivada lateral derecha ni tampoco una derivada lateral izquierda en $x = 0$.

A.2 Comentarios sobre la derivabilidad de funciones

La derivada de una función que es continua y posee derivada en cada punto de un intervalo puede no ser continua.

Como un ejemplo sencillo considérese la función dada por

$$y = f(x) = x^2 \operatorname{sen} \frac{1}{x} \qquad \text{para } x \neq 0$$

y

$$f(0) = 0.$$

Esta función está definida y es continua en todas partes. Para todos los valores de x diferentes de cero la derivada está dada por la expresión

$$f'(x) = -x^2 \left(\cos \frac{1}{x}\right) \frac{1}{x^2} + 2x \operatorname{sen} \frac{1}{x} = -\cos \frac{1}{x} + 2x \operatorname{sen} \frac{1}{x}.$$

Cuando x tiende a cero $f'(x)$ no posee límite. Si, por otra parte, se forma el cociente incremental $[f(h) - f(0)]/h = (h^2 \operatorname{sen} 1/h)/h = h \operatorname{sen} 1/h$, se ve de inmediato que éste tiende a cero cuando h lo hace. La derivada, por lo tanto, existe para $x = 0$ y posee el valor 0.

Para entender intuitivamente la razón de este comportamiento paradójico se representa la función gráficamente (ver Fig. 3.26). Esta oscila entre las curvas $y = x^2$ e $y = -x^2$ a las cuales toca alternadamente. Así, las razones de las alturas de las crestas de onda de la curva y sus distancias al

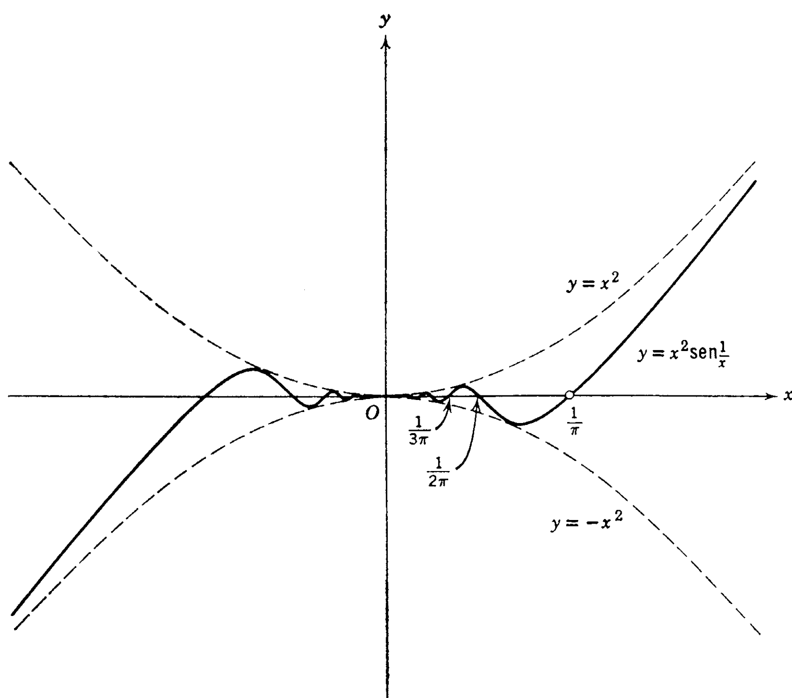


Figura 3.26.

origen se tornan gradualmente más pequeñas. Sin embargo, estas ondas no se tornan más planas pues su pendiente, dada por la derivada

$$f'(x) = 2x \operatorname{sen} 1/x - \cos 1/x,$$

es igual a -1 en los puntos $x = 1/2\, n\pi$, donde $\cos 1/x = 1$, e igual a $+1$ en los puntos $x = 1/(2n+1)\pi$, donde $\cos 1/x = -1$.

En contraste con la posibilidad ilustrada aquí (de que una derivada existe en todas partes y sin embargo no es continua), se establece el siguiente sencillo teorema, el cual ilumina toda una serie de ejemplos y discusiones anteriores.

TEOREMA. *Si se sabe que en una vecindad de un punto $x = a$ la función $f(x)$ es continua, que para $x \neq a$ también posee derivada $f'(x)$ y si, además, la ecuación $\lim_{x \to a} f'(x) = b$ se cumple, entonces la derivada $f'(x)$ existe también en el punto a y $f'(a) = b$.*

DEMOSTRACIÓN. La demostración se obtiene inmediatamente a partir del teorema del valor medio. Pues se tiene $[f(a+h) - f(a)]/h = f'(\xi)$, donde ξ es un valor intermedio entre a y $a + h$. Si h tiende ahora a cero, por hipótesis $f'(\xi)$ tiende a b, y la tesis surge de inmediato.

Un teorema análogo puede ser probado de un modo similar: Si la función $f(x)$ es continua en $a \leq x \leq b$ y para $a < x < b$ posee una derivada que crece más allá de toda cota conforme x tiende a a, el cociente incremental derecho $[f(a+h) - f(a)]/h$ crece también más allá de toda cota conforme h tiende a cero, de modo que no existe derivada lateral derecha para $x = a$. El significado geométrico de este aserto es que en el punto con coordenadas (finitas) $[a, f(a)]$ la curva posee una tangente vertical.

Parte B Técnicas de integración

Funciones explícitas

Una amplia clase de funciones puede ser construida a partir de las funciones elementales [1] aplicando repetidamente operaciones racionales, esto es, la adición, multiplicación, división, y, además, mediante las operaciones de formar funciones inversas y funciones compuestas. Las funciones así descriptas constituyen la clase de *funciones "explícitas"* o "expresiones exactas" finitas.[2] Como un resultado de la **Parte A** de este capítulo se establece la propiedad bastante general:

Toda función explícita puede derivarse y su derivada es también una función explícita.

Se ha logrado así un dominio bastante completo de la operación o "algoritmo" de la diferenciación. No obstante, el proceso inverso, el de integra-

---

[1] Se deberá destacar que la distinción entre funciones "elementales" y "explícitas" y otras es, de por sí, algo arbitraria. Para nosotros el término función "elemental" incluye justamente las funciones racionales, las funciones trigonométricas y las exponenciales, y sus inversas.

[2] Este nombre indica que se encontrarán muchas otras funciones que no pueden ser representadas de esta forma pero que pueden ser construidas por medio de procesos de límite tales como las series infinitas.

ción, es, hablando generalmente, más importante y representa el mayor problema. Hasta cierto grado el problema es abordado y resuelto mediante el teorema fundamental del cálculo: A cada fórmula de diferenciación $F'(x) = f(x)$ le corresponde una fórmula equivalente para las funciones primitivas $F(x)$ de $f(x)$, o sea, la integral:

$$\int f(x)\,dx = F(x).$$

(Más precisamente, se tiene $F(x) = \int_a^x f(u)\,du +$ constante). Así, mientras más fórmulas explícitas de diferenciación sean deducidas, más funciones explícitas podrán ser integradas en términos de funciones explícitas. Una primera tabla de integrales se da en la p. 284. En principio, no sería difícil, aunque sí impráctico y confuso, extender mucho más dicha tabla.

En las primeras fases del desarrollo del cálculo muchos matemáticos trataron de encontrar, en forma explícita o exacta, la función integral o primitiva para toda función dada explícitamente.

Pasó algún tiempo hasta que se comprendió que, en principio, este problema no puede ser resuelto. Por el contrario, para algunos integrandos bastante elementales la integral no puede ser expresada en términos de funciones elementales (véase Sec. 3.14*a*). Así, la necesidad de estudiar nuevos tipos de funciones generadas mediante procesos de integración a partir de funciones elementales se convirtió en un estímulo importante para el desarrollo del análisis. Sin embargo, el deseo de integrar—cuando es factible—funciones explícitas dadas explícitamente, sin meterse en los enredos desesperados de la consulta tediosa de tablas o de los cálculos numéricos, ha conducido a algunos recursos sencillos que proporcionan una cierta flexibilidad para transformar integrales dadas. En efecto, estos recursos permiten efectuar la integración mediante la reducción a alguna de las integrales elementales en la Tabla de Integrales.

La sección 3.9 será dedicada al desarrollo de tales útiles recursos. En relación con esto, el principiante deberá estar precavido en contra de memorizar simplemente las muchas fórmulas obtenidas usando estos recursos técnicos. En vez de ello, el estudiante deberá dirigir su esfuerzo hacia la obtención de una comprensión clara de los *métodos* de integración y el aprendizaje de cómo aplicarlos. Más aún, deberá recordar que aun cuando la integración mediante estos recursos sea imposible, la *integral sí existe* (al menos para todas las funciones continuas), y puede ser realmente calculada con un grado tan alto de precisión como se desee por medio de los métodos numéricos que se desarrollarán posteriormente (sección 6.1).

La Parte C de este capítulo se ocupará de extender los conceptos de integración e integral, en forma separada del problema de la técnica de la integración.

Tabla de integrales elementales

	$F'(x) = f(x)$	$F(x) = \int f(x)\,dx$
1.	$x^a \quad (a \neq -1)$	$\dfrac{x^{a+1}}{a+1}$
2.	$\dfrac{1}{x}$	$\log\|x\|$
3.	e^x	e^x
4.	$a^x \quad (a \neq 1)$	$\dfrac{a^x}{\log a}$
5.	$\operatorname{sen} x$	$-\cos x$
6.	$\cos x$	$\operatorname{sen} x$
7.	$\dfrac{1}{\operatorname{sen}^2 x} \quad (= \operatorname{cosec}^2 x)$	$-\cot x$
8.	$\dfrac{1}{\cos^2 x} \quad (= \sec^2 x)$	$\tan x$
9.	$\operatorname{senh} x$	$\cosh x$
10.	$\cosh x$	$\operatorname{senh} x$
11.	$\dfrac{1}{\operatorname{senh}^2 x} \quad (= \operatorname{cosech}^2 x)$	$-\coth x$
12.	$\dfrac{1}{\cosh^2 x} \quad (= \operatorname{sech}^2 x)$	$\tanh x$
13.	$\dfrac{1}{\sqrt{1-x^2}} \quad (\|x\| < 1)$	$\begin{cases} \operatorname{arc\,sen} x \\ -\operatorname{arc\,cos} x \end{cases}$
14.	$\dfrac{1}{1+x^2}$	$\begin{cases} \operatorname{arc\,tan} x \\ -\operatorname{arc\,cot} x \end{cases}$
15.	$\dfrac{1}{\sqrt{1+x^2}}$	$\operatorname{arc\,senh} x = \log(x + \sqrt{1+x^2})$
16.	$\dfrac{1}{\pm\sqrt{x^2-1}} \quad (\|x\| > 1)$	$\operatorname{arc\,cosh} x = \log(x \pm \sqrt{x^2-1})$
17.	$\dfrac{1}{1-x^2} \quad \|x\| < 1$	$\operatorname{arc\,tanh} x = \dfrac{1}{2}\log\dfrac{1+x}{1-x}$
	$\dfrac{1}{1-x^2} \quad \|x\| > 1$	$\operatorname{arc\,coth} x = \dfrac{1}{2}\log\dfrac{x+1}{x-1}$

3.8 Tabla de integrales elementales

A cada una de las fórmulas de derivación demostradas anteriormente le corresponde una fórmula equivalente de integración. Puesto que estas integrales elementales son utilizadas una y otra vez como materiales para el arte de la integración, se han reunido en una tabla. La columna derecha contiene un número de funciones elementales y la columna izquierda las correspondientes derivadas. Si se lee la tabla de izquierda a derecha, se obtiene en la columna derecha una integral indefinida de la función que aparece en la columna izquierda.

Se le recuerdan también al lector los teoremas fundamentales del cálculo diferencial e integral, demostrados en la sección 2.9, en particular, el hecho de que cualquier integral definida es obtenida a partir de la integral indefinida $F(x)$ mediante la fórmula [1]

$$\int_a^b f(x)\, dx = F(x) \Big|_a^b = F(b) - F(a).$$

En las secciones siguientes se intentará reducir el cálculo de integrales de funciones dadas de una u otra manera al de las integrales elementales coleccionadas en nuestra tabla. Aparte de algunos artificios especiales que son aprendidos solamente por experiencia, esta reducción está basada esencialmente en dos métodos útiles: "substitución" e "integración por partes". Cada uno de estos métodos permite transformar de muchas maneras una integral dada; y el objeto de tales transformaciones en su mayor parte es el de reducir la integral dada, en un paso o en una sucesión de pasos, a una o más de las fórmulas elementales de integración dadas arriba.

3.9 El método de substitución

Integración de funciones compuestas

El primero de estos métodos consiste en la *introducción de una nueva variable* (esto es, el método de *substitución* o *transformación).* Su objeto es el de reducir la integración de funciones compuestas (tales como funciones de $x - c$ o de $ax + b$) a la de funciones más sencillas.

a. La fórmula de substitución. Integral de una función compuesta

La regla para integrar funciones compuestas se obtiene a partir de la correspondiente regla de la cadena para derivar. Para una función compuesta $G(u) = F[\phi(u)]$ se tiene (véase p. 239)

---

[1] En este capítulo no se discutirá el problema algo diferente de calcular integrales definidas especiales sin encontrar primero una función primitiva general.

$$(16) \qquad \frac{dG(u)}{du} = \frac{dF[\phi(u)]}{du} = F'[\phi(u)]\phi'(u).$$

Para la validez de esta fórmula es suficiente que las funciones $x = \phi(u)$ y $F(x)$ sean continuamente derivables respecto a sus argumentos u, x, respectivamente, y que $F(x)$ esté definida para los valores de x tomados por la función $x = \phi(u)$ (esto es, el rango de la función ϕ debe pertenecer al dominio de F). Integrando la fórmula entre los límites $u = \alpha$ y $u = \beta$ se encuentra

$$(17) \qquad G(\beta) - G(\alpha) = F[\phi(\beta)] - F[\phi(\alpha)] = \int_\alpha^\beta F'[\phi(u)]\phi'(u)\, du.$$

Si aquí se hace

$$\phi(\beta) = b, \qquad \phi(\alpha) = a,$$

se tiene

$$F[\phi(\beta)] - F[\phi(\alpha)] = F(b) - F(a) = \int_a^b F'(x)\, dx.$$

Haciendo $F'(x) = f(x)$ se obtiene la *fórmula fundamental de substitución*

$$(18) \qquad \int_a^b f(x)\, dx = \int_\alpha^\beta f[\phi(u)]\phi'(u)\, du, \qquad x = \phi(u),$$

o sea, escrita sugestivamente en la notación de Leibnitz con la diferencial $d\phi = \phi'(x)\, dx$,

$$(18a) \qquad \int f(x)\, dx = \int f(\phi)\, d\phi.$$

Aquí $x = \phi(u)$ puede ser cualquier función que esté definida y posea una derivada continua en el intervalo J con puntos extremos α y β; y dicha función transforma estos puntos extremos en $x = a$ y $x = b$ respectivamente. Se supone que la función $f(x)$ es continua en un intervalo I que contiene las imágenes de todos los puntos de J bajo la transformación ϕ. Por $F(x)$ puede tomarse *cualquier función primitiva de* $f(x)$.

Como se habrá notado, la regla de la substitución (18) no requiere que la transformación $x = \phi(u)$ transforme puntos entre α y β solamente en puntos entre a y b, o sea, que valores diferentes u sean transformados en diferentes x; todo lo que importa es que α y β sean transformados en a y b y que $f(x)$ esté definida para los valores x tomados por $\phi(u)$ para u entre α y β.

En términos de integrales indefinidas la regla de la substitución adquiere la forma

$$(19) \qquad G(u) = \int f[\phi(u)]\phi'(u)\, du = \int f(x)\, dx = F(x) = F[\phi(u)].$$

Los símbolos diferenciales

$$\phi'(u)\,du = \frac{dx}{du}du \qquad \text{y} \qquad dx$$

se vuelven idénticos si formalmente se cancelan los símbolos du en el numerador y el denominador.

Ejemplos. Se aplica la fórmula (18) al integrando $f(x) = 1/x$ y se hace la substitución $x = \phi(u)$, suponiendo $\phi(u) \neq 0$ en el intervalo considerado; entonces

$$\int \frac{\phi'(u)}{\phi(u)}\,du = \int \frac{dx}{x} = \log|x| = \log|\phi(u)|,$$

o bien, cambiando el nombre de la variable u nuevamente por x,

$$\int \frac{\phi'(x)}{\phi(x)}\,dx = \log|\phi(x)|. \tag{20}$$

Si en esta importante fórmula se substituyen funciones particulares, tales como $\phi(x) = \log x$, $\phi(x) = \operatorname{sen} x$, o $\phi(x) = \cos x$, se obtiene[1]

$$\int \frac{dx}{x\log x} = \log|\log x|,$$

(21)

$$\int \cot x\,dx = \log|\operatorname{sen} x|, \qquad \int \tan x\,dx = -\log|\cos x|.$$

Otros ejemplos

1) $$\int \phi(u)\phi'(u)\,du = \int x\,dx = \tfrac{1}{2}x^2 = \tfrac{1}{2}[\phi(u)]^2,$$

donde $f(x) = x$. De aquí se obtiene para $\phi(u) = \log u$:

$$\int \frac{\log u}{u}\,du = \tfrac{1}{2}(\log u)^2. \tag{22}$$

2) Finalmente se considera

$$\int \operatorname{sen}^n u \cos u\,du.$$

Aquí $x = \operatorname{sen} u = \phi(u)$, y, por lo tanto,

$$\int \operatorname{sen}^n u \cos u\,du = \int x^n\,dx = \frac{x^{n+1}}{n+1} = \frac{\operatorname{sen}^{n+1} u}{n+1}.$$

[1] Estas y las siguientes fórmulas son verificadas fácilmente mostrando que la derivada del resultado da otra vez el integrando.

3) La misma substitución $x = \operatorname{sen} u$ da, para cualquier función $f(x)$ continua en el intervalo $-1 \leq x \leq 1$,

$$\int_{\alpha}^{\beta} f(\operatorname{sen} u) \cos u \, du = \int_{\operatorname{sen} \alpha}^{\operatorname{sen} \beta} f(x) \, dx.$$

Tomando aquí $\alpha = 0$ y $\beta = 2\pi$, resulta un ejemplo para aplicar la fórmula de substitución a un caso en que la función transformación $x = \phi(u) = \operatorname{sen} u = x$ no es monótona en el intervalo $\alpha \leq u \leq \beta$. Se encuentra

$$\int_{0}^{2\pi} f(\operatorname{sen} u) \cos u \, du = \int_{0}^{0} f(x) \, dx = 0.$$

Otras formas de la regla

En muchas aplicaciones la integral que ha de evaluarse está dada en la forma

$$F(u) = \int h[\phi(u)] \, du$$

en la cual el integrando aparece como la función compuesta $h[\phi(u)]$ sin el factor adicional $\phi'(u)$. Puede aplicarse la regla de substitución (18) si se logra escribir el integrando $h[\phi(u)]$ en la forma $f[\phi(u)]\phi'(u)$. Esto puede ser logrado siempre bajo la hipótesis de que la función $x = \phi(u)$ posee una derivada continua $\phi'(u)$ *que no se anula.* Porque entonces existe una función inversa $u = \psi(x)$ con derivada continua $du/dx = \psi'(x) = 1/\phi'(u)$. Tomando por $f(x)$ la función $h(x)\psi'(x)$, se tiene así $h[\phi(u)] = f[\phi(u)]/\psi'(x) = f[\phi(u)]\phi'(u)$ y se obtiene, a partir de la regla de substitución,

$$(23) \qquad \int h[\phi(u)] \, du = \int f[\phi(u)]\phi'(u) \, du = \int f(x) \, dx$$
$$= \int h(x)\psi'(x) \, dx = \int h(x) \frac{du}{dx} dx.$$

La hipótesis $\phi'(u) \neq 0$ ha sido introducida para evitar que la expresión dx/du en la fórmula (23) se torne infinita.

El principiante jamás deberá olvidar que al substituir $\psi(x)$ por u en una integral no se debe expresar simplemente la antigua variable x en términos de la nueva, u, y después integrar con respecto a esta nueva variable; en vez de esto, *antes* de integrar se debe multiplicar el integrando por la derivada de la variable original x con respecto a la nueva variable u. Esto, por supuesto, es sugerido por la notación de Leibnitz $h \, dx = h \dfrac{dx}{du} du$. En la integral definida

$$\int_{a}^{b} h[\psi(x)] \, dx = \int_{\alpha}^{\beta} h(u)\phi'(u) \, du$$

no debe olvidarse cambiar los límites a, b, para x, por los límites correspondientes $\alpha = \psi(a)$ y $\beta = \psi(b)$ para la variable u.

Ejemplos. Para calcular $\int \operatorname{sen} 2x\, dx$ se toma $u = \psi(x) = 2x$ y $h(u) = \operatorname{sen} u$. Se tiene entonces

$$\frac{du}{dx} = \psi'(x) = 2, \quad \frac{dx}{du} = \frac{1}{2}.$$

Si se introduce ahora $u = 2x$ en la integral como la nueva variable, entonces aquélla se transforma, *no* en $\int \operatorname{sen} u\, du$, sino en

$$\frac{1}{2}\int \operatorname{sen} u\, du = -\frac{1}{2}\cos u = -\frac{1}{2}\cos 2x;$$

esto puede, por supuesto, verificarse de inmediato derivando el lado derecho de la ecuación.

Si se integra con respecto a x entre los límites cero y $\pi/4$, los correspondientes límites para $u = 2x$ son cero y $\pi/2$, y se obtiene

$$\int_0^{\pi/4} \operatorname{sen} 2x\, dx = \frac{1}{2}\int_0^{\pi/2} \operatorname{sen} u\, du = -\frac{1}{2}\cos u \Big|_4^{\pi/2} = \frac{1}{2}$$

Otro ejemplo sencillo es la integral $\int_1^4 \frac{dx}{\sqrt{x}}$. Aquí se toma $u = \psi(x) = \sqrt{x}$, de donde $x = \phi(u) = u^2$. Puesto que $\phi'(u) = 2u$, se tiene

$$\int_1^4 \frac{dx}{\sqrt{x}} = \int_1^2 2\frac{u\, du}{u} = 2\int_1^2 du = 2.$$

Como otro ejemplo, se considera la integral de $\operatorname{sen} 1/x$ para el intervalo $\frac{1}{2} \leq x \leq 1$. Se tiene para $u = 1/x$ o bien $x = 1/u$, $dx = -du/u^2$; y, por lo tanto,

$$\int_{1/2}^1 \operatorname{sen}\frac{1}{x}\, dx = -\int_2^1 \frac{\operatorname{sen} u}{u^2}\, du = \int_1^2 \frac{\operatorname{sen} u}{u^2}\, du.$$

*b. Una deducción alternativa de la fórmula de substitución

La fórmula de integración (17), con una ligera variante en la notación, puede ser intepretada también de un modo directo basado en el significado de la integral definida como el límite de una suma, en vez de ser deducida a partir de la regla de la cadena de la derivación.[1] Para calcular la integral

$$\int_a^b h[\psi(x)]\, dx$$

(para el caso $a < b$), se parte con una subdivisión arbitraria del intervalo $a \leq x \leq b$ y se hace entonces la subdivisión cada vez más fina. Estas subdivisiones se inician de la siguiente manera. Si se supone que la función $u = \psi(x)$ es monótona creciente, existe una correspondencia biunívoca entre el intervalo $a \leq x \leq b$ sobre el eje x y un intervalo $\alpha \leq u \leq \beta$ de los

[1] El resultado obtenido de esta manera está restringido nuevamente a substituciones monótonas, y así es menos general que la fórmula (18) proporcionada por la regla de la cadena (pág. 286).

valores de $u = \psi(x)$, donde $\alpha = \psi(a)$ y $\beta = \psi(b)$. Se divide este intervalo x en n partes de longitud[1] Δx; existe *una correspondiente* subdivisión del intervalo u en subintervalos que, en general, no son todos de la misma longitud. Los puntos de división del intervalo x se denotan por

$$x_0 = a, x_1, x_2, \ldots, x_n = b,$$

y las longitudes de las correspondientes celdas u por

$$\Delta u_1, \Delta u_2, \ldots, \Delta u_n.$$

La integral que se está considerando es el límite de la suma

$$\sum_{\nu=1}^{n} h\{\psi(\xi_\nu)\} \Delta x,$$

donde el valor ξ_ν es seleccionado arbitrariamente en el ν-ésimo subintervalo de la subdivisión x. Esta suma se escribe ahora en la forma $\sum_{\nu=1}^{n} h(v_\nu) \frac{\Delta x}{\Delta u_\nu} \Delta u_\nu$, donde $v_\nu = \psi(\xi_\nu)$. Por el teorema del valor medio del cálculo diferencial, es $\Delta x/\Delta u_\nu = \phi'(\eta_\nu)$, donde η_ν es un valor intermedio elegido adecuadamente para la variable u en el ν-ésimo subintervalo de la subdivisión u, y $x = \phi(u)$ denota la función inversa de $u = \psi(x)$. Si se selecciona ahora el valor ξ_ν de tal manera que v_ν y η_ν coincidan, esto es, $\eta_\nu = \psi(\xi_\nu)$, $\xi_\nu = \phi(\eta_\nu)$, entonces la suma toma la forma

$$\sum_{\nu=1}^{n} h(\eta_\nu)\phi'(\eta_\nu) \Delta u_\nu.$$

Si ahora aquí se pasa al límite haciendo $n \to \infty$,[2] se obtiene la expresión

$$\int_\alpha^\beta h(u) \frac{dx}{du} du$$

como el valor límite, esto es, como el valor de la integral que se está considerando, de acuerdo con la fórmula (23) dada anteriormente.

Así se llega al siguiente resultado.

TEOREMA. *Sea $h(u)$ una función continua de u en el intervalo $\alpha \leq u \leq \beta$. Si la función $u = \psi(x)$ es continua y monótona y posee una derivada continua du/dx que no se anula en $a \leq x \leq b$, y si, además, $\psi(a) = \alpha$ y $\psi(b) = \beta$*, entonces

$$\int_a^b h\{\psi(x)\} dx = \int_a^b h(u) dx = \int_\alpha^\beta h(u) \frac{dx}{du} du.$$

---

[1] La hipótesis de que las longitudes de estos subintervalos son iguales no es de ninguna manera esencial para la demostración.

[2] Este límite existe (para $\Delta x \to 0$) y es la integral, puesto que, de acuerdo con la continuidad uniforme de $u = \psi(x)$, la mayor de las longitudes Δu_ν tiende a cero con Δx.

Esta deducción exhibe el mérito sugestivo de la notación de Leibnitz. Para llevar a cabo la substitución $u = \psi(x)$ se requiere escribir solamente $(dx/du)du$ en vez de dx, cambiando los límites de los valores originales de x a los correspondientes valores de u.

c. Ejemplos. Fórmulas de integración

Con la ayuda de la regla de substitución puede evaluarse en muchos casos una integral dada, $\int f(x)\,dx$, si se le reduce por medio de una substitución adecuada, $x = \phi(u)$, a una de las integrales fundamentales de nuestra tabla. Si tales substituciones existen y cómo encontrarlas son problemas a los cuales no se les puede dar una respuesta general; y ésta es, más bien, una cuestión en la cual la práctica y el ingenio, en contraste con los métodos sistemáticos, desempeñan el papel más importante.

Como un ejemplo se evalúa la integral $\int \frac{dx}{\sqrt{a^2 - x^2}}$ por medio de la substitución [1] $x = \phi(u) = au$, $u = \psi(x) = x/a$, $dx = a\,du$, mediante la cual, usando el Nº 13 de nuestra tabla, se obtiene

(24)
$$\int \frac{dx}{\sqrt{a^2 - x^2}} = \int \frac{a\,du}{a\sqrt{1 - u^2}} = \operatorname{arc\,sen} u = \operatorname{arc\,sen} \frac{x}{a}, \qquad \text{para } |x| < |a|.$$

Mediante la misma substitución se obtienen

$$(25) \qquad \int \frac{dx}{a^2 + x^2} = \int \frac{a\,du}{a^2(1 + u^2)} = \frac{1}{a} \operatorname{arc\,tan} u = \frac{1}{a} \operatorname{arc\,tan} \frac{x}{a},$$

$$(26) \qquad \int \frac{dx}{\sqrt{a^2 + x^2}} = \operatorname{ar\,senh} \frac{x}{a},$$

$$(27) \qquad \int \frac{dx}{\sqrt{x^2 - a^2}} = \operatorname{ar\,cosh} \frac{x}{a}, \qquad \text{para } |x| > |a|,$$

$$(28) \qquad \int \frac{dx}{a^2 - x^2} = \begin{cases} \dfrac{1}{a} \operatorname{ar\,tanh} \dfrac{x}{a} & \text{para } |x| < |a|, \\[2ex] \dfrac{1}{a} \operatorname{ar\,coth} \dfrac{x}{a} & \text{para } |x| > |a|, \end{cases}$$

fórmulas que aparecen muy frecuentemente y que pueden ser fácilmente verificadas derivando el lado derecho de cada ecuación.

[1] Por brevedad nos tomamos nuevamente la libertad de escribir los símbolos dx y du separadamente, esto es $dx = \phi'(u)\,du$ en vez de $dx/du = \phi'(u)$ (Ver. p. 201).

3.10 Otros ejemplos del método de substitución

En esta sección se recopila un número de ejemplos que el lector puede considerar cuidadosamente para ejercitación.

Mediante la substitución $u = 1 \pm x^2$, $du = \pm 2x\,dx$, se deduce que

$$\int \frac{x\,dx}{\sqrt{1 \pm x^2}} = \pm\sqrt{1 \pm x^2}, \tag{29}$$

$$\int \frac{x\,dx}{1 \pm x^2} = \pm\tfrac{1}{2}\log|1 \pm x^2|. \tag{30}$$

En estas fórmulas debe tomarse en los tres lugares el mismo signo, ya sea el positivo o bien el negativo.

Mediante la substitución $u = ax + b$, $du = a\,dx$ $(a \neq 0)$, se obtiene

$$\int \frac{dx}{ax + b} = \frac{1}{a}\log|ax + b|, \tag{31}$$

$$\int (ax + b)^{\alpha}\,dx = \frac{1}{a(\alpha + 1)}(ax + b)^{\alpha+1} \qquad (\alpha \neq -1), \tag{32}$$

$$\int \operatorname{sen}(ax + b)\,dx = -\frac{1}{a}\cos(ax + b); \tag{33}$$

y análogamente, por medio de la substitución $u = \cos x$, $du = -\operatorname{sen} x\,dx$, se obtiene

$$\int \tan x\,dx = -\log|\cos x|; \tag{34}$$

y, por medio de la substitución $u = \operatorname{sen} x$, $du = \cos x\,dx$, se tiene

$$\int \cot x\,dx = \log|\operatorname{sen} x| \tag{35}$$

[Ver (21) p. 287]. Utilizando las substituciones análogas $u = \cosh x$ (que da $du = \operatorname{senh} x\,dx$) y $u = \operatorname{senh} x$ $(du = \cosh x\,dx)$, se obtienen las fórmulas

$$\int \tanh x\,dx = \log\cosh x, \tag{36}$$

$$\int \coth x\,dx = \log|\operatorname{senh} x|. \tag{37}$$

Por medio de la substitución $u = (a/b)\tan x$, $du = (a/b)\sec^2 x\,dx$, se llega a las dos fórmulas

$$\int \frac{dx}{a^2\operatorname{sen}^2 x + b^2\cos^2 x} = \frac{1}{b^2}\int \frac{1}{(a^2/b^2)\tan^2 x + 1}\,\frac{dx}{\cos^2 x} \tag{38}$$

$$= \begin{cases} \dfrac{1}{ab}\operatorname{arc\,tan}\left(\dfrac{a}{b}\tan x\right) \\[2ex] \dfrac{1}{ab}\operatorname{arc\,cot}\left(\dfrac{a}{b}\tan x\right) \end{cases},$$

y

(39) $$\int \frac{dx}{a^2 \operatorname{sen}^2 x - b^2 \cos^2 x} = \begin{cases} -\dfrac{1}{ab} \operatorname{ar\,tanh}\left(\dfrac{a}{b}\tan x\right) \\ -\dfrac{1}{ab} \operatorname{ar\,coth}\left(\dfrac{a}{b}\tan x\right) \end{cases}$$

Se evalúa la integral

$$\int \frac{dx}{\operatorname{sen} x}$$

escribiendo $\operatorname{sen} x = 2 \operatorname{sen} (x/2) \cos (x/2) = 2 \tan (x/2) \cos^2 (x/2)$ y haciendo $u = \tan (x/2)$, de modo que $du = \frac{1}{2} \sec^2 (x/2)\, dx$; la integral se vuelve entonces

(40) $$\int \frac{dx}{\operatorname{sen} x} = \int \frac{du}{u} = \log \left| \tan \frac{x}{2} \right| .$$

Si se reemplaza x por $x + \pi/2$, esta fórmula se convierte en

(41) $$\int \frac{dx}{\cos x} = \log \left| \tan \left(\frac{x}{2} + \frac{\pi}{4} \right) \right| .$$

La substitución $u = 2x$ conduce, si se aplican también las conocidas fórmulas trigonométricas $2 \cos^2 x = 1 + \cos 2x$ y $2 \operatorname{sen}^2 x = 1 - \cos 2x$, a las fórmulas frecuentemente utilizadas

(42) $$\int \cos^2 x \, dx = \tfrac{1}{2}(x + \operatorname{sen} x \cos x)$$

y

(43) $$\int \operatorname{sen}^2 x \, dx = \tfrac{1}{2}(x - \operatorname{sen} x \cos x).$$

Mediante la substitución $x = \cos u$, equivalente a $u = \operatorname{arc\,cos} x$, o, más generalmente, $x = a \cos u$ $(a \neq 0)$, podemos reducir

$$\int \sqrt{1 - x^2}\, dx \quad \text{y} \quad \int \sqrt{a^2 - x^2}\, dx$$

respectivamente a estas fórmulas. Se obtiene así

(44) $$\int \sqrt{a^2 - x^2}\, dx = -\frac{a^2}{2} \operatorname{arc\,cos} \frac{x}{a} + \frac{x}{2} \sqrt{a^2 - x^2}.$$

Análogamente, mediante la substitución $x = a \cosh u$ se obtiene la fórmula

(45) $$\int \sqrt{x^2 - a^2}\, dx = -\frac{a^2}{2} \operatorname{ar\,cosh} \frac{x}{a} + \frac{x}{2} \sqrt{x^2 - a^2},$$

y mediante la substitución $x = a \operatorname{senh} u$ se tiene

(46) $$\int \sqrt{a^2 + x^2}\, dx = \frac{a^2}{2} \operatorname{ar\,senh} \frac{x}{a} + \frac{x}{2} \sqrt{a^2 + x^2}.$$

La substitución $u = a/x$, $dx = -(a/u^2)\,du$, conduce a las fórmulas

$$(47)\qquad \int \frac{dx}{x\sqrt{x^2 - a^2}} = -\frac{1}{a}\operatorname{arc\,sen}\frac{a}{x},$$

$$(48)\qquad \int \frac{dx}{x\sqrt{x^2 + a^2}} = -\frac{1}{a}\operatorname{ar\,senh}\frac{a}{x};$$

$$(49)\qquad \int \frac{dx}{x\sqrt{a^2 - x^2}} = -\frac{1}{a}\operatorname{ar\,cosh}\frac{a}{x}.$$

Finalmente, se consideran las tres integrales

$$\int \operatorname{sen} mx \operatorname{sen} nx\, dx, \qquad \int \operatorname{sen} mx \cos nx\, dx, \qquad \int \cos mx \cos nx\, dx,$$

donde m y n son enteros positivos. Mediante bien conocidas fórmulas trigonométricas puede dividirse cada una de estas integrales en dos partes, escribiendo

$$\operatorname{sen} mx \operatorname{sen} nx = \tfrac{1}{2}[\cos(m - n)x - \cos(m + n)x],$$

$$\operatorname{sen} mx \cos nx = \tfrac{1}{2}[\operatorname{sen}(m + n)x + \operatorname{sen}(m - n)x],$$

$$\cos mx \cos nx = \tfrac{1}{2}[\cos(m + n)x + \cos(m - n)x].$$

Si se hace ahora uso de las substituciones $u = (m + n)x$ y $u = (m - n)x$ respectivamente, se obtiene directamente el siguiente sistema de fórmulas:

(50)

$$\int \operatorname{sen} mx \operatorname{sen} nx\, dx = \begin{cases} \dfrac{1}{2}\left\{\dfrac{\operatorname{sen}(m - n)x}{m - n} - \dfrac{\operatorname{sen}(m + n)x}{m + n}\right\} & \text{si } m \neq n, \\[2ex] \dfrac{1}{2}\left(x - \dfrac{\operatorname{sen} 2mx}{2m}\right) & \text{si } m = n; \end{cases}$$

(51)

$$\int \operatorname{sen} mx \cos nx\, dx = \begin{cases} -\dfrac{1}{2}\left\{\dfrac{\cos(m + n)x}{m + n} + \dfrac{\cos(m - n)x}{m - n}\right\} & \text{si } m \neq n \\[2ex] -\dfrac{1}{2}\left(\dfrac{\cos 2mx}{2m}\right) & \text{si } m = n; \end{cases}$$

(52)

$$\int \cos mx \cos nx\, dx = \begin{cases} \dfrac{1}{2}\left\{\dfrac{\operatorname{sen}(m + n)x}{m + n} + \dfrac{\operatorname{sen}(m - n)x}{m - n}\right\} & \text{si } m \neq n, \\[2ex] \dfrac{1}{2}\left(\dfrac{\operatorname{sen} 2mx}{2m} + x\right) & \text{si } m = n. \end{cases}$$

Si, en particular, se integra de $-\pi$ a $+\pi$, se obtienen de estas fórmulas las relaciones sumamente importantes

$$\int_{-\pi}^{+\pi} \operatorname{sen} mx \operatorname{sen} nx \, dx = \begin{cases} 0 & \text{si } m \neq n, \\ \pi & \text{si } m = n, \end{cases}$$

(53) $$\int_{-\pi}^{+\pi} \operatorname{sen} mx \cos nx \, dx = 0,$$

$$\int_{-\pi}^{+\pi} \cos mx \cos nx \, dx = \begin{cases} 0 & \text{si } m \neq n, \\ \pi & \text{si } m = n. \end{cases}$$

Estas son las *relaciones de ortogonalidad* de las funciones trigonométricas, las cuales se encontrarán nuevamente en la sección 8.4e.

3.11 Integración por partes

a. Fórmula general

El segundo método ampliamente utilizado para tratar problemas de integración expresa en forma integral la regla para derivar un producto,

$$(fg)' = f'g + fg'.$$

La correspondiente fórmula integral es (Ver. p. 210)

$$f(x)g(x) = \int g(x)f'(x)\, dx + \int f(x)g'(x)\, dx,$$

o bien,

(54) $$f(x)g'(x)\, dx = f(x)g(x) - \int g(x)f'(x)\, dx.$$

Utilizando la notación diferencial de Leibnitz, esta expresión se convierte en

(54a) $$\int f\, dg = fg - \int g\, df.$$

Esta fórmula será mencionada como la *fórmula de integración por partes.* Ella reduce el cálculo de una integral al cálculo de otra integral. Puesto que un integrando dado puede ser interpretado como un producto $f(x)g'(x)$ de muchas maneras diferentes, esta fórmula proporciona una herramienta efectiva para la transformación de integrales.

Escrita como fórmula de *integración definida,* la fórmula para integración por partes es

(54b) $$\int_a^b f(x)g'(x)\, dx = f(x)g(x)\Big|_a^b - \int_a^b g(x)f'(x)\, dx$$

$$= f(b)g(b) - f(a)g(a) - \int_a^b g(x)f'(x)\, dx.$$

Esta se obtiene, ya sea directamente integrando la fórmula de la derivada de un producto entre los límites a y b, o bien formando la diferencia de la fórmula (54) en los puntos b y a.

Puede darse una interpretación geométrica sencilla de la fórmula (54b): Supóngase que $y = f(x)$ y $z = g(x)$ son funciones monótonas y que $f(a) = A$, $f(b) = B$, $g(a) = \alpha$, $g(b) = \beta$; puede entonces formarse la inversa de la primera función y substituir en la segunda ecuación, obteniéndose así z como función de y. Se supone que esta función es monótona creciente. Puesto que $dy = f'(x)\,dx$ y $dz = g'(x)\,dx$, la fórmula para integración por partes puede ser escrita [Ver la regla de substitución (18), p. 286] como

$$\int_{\alpha}^{\beta} y\,dz + \int_{A}^{B} z\,dy = B\beta - A\alpha,$$

en concordancia con la relación que se pone de manifiesto mediante la Fig. 3.27:

área $NQLK$ + área $PMLQ$ = área $OMLK$ − área $OPQN$.

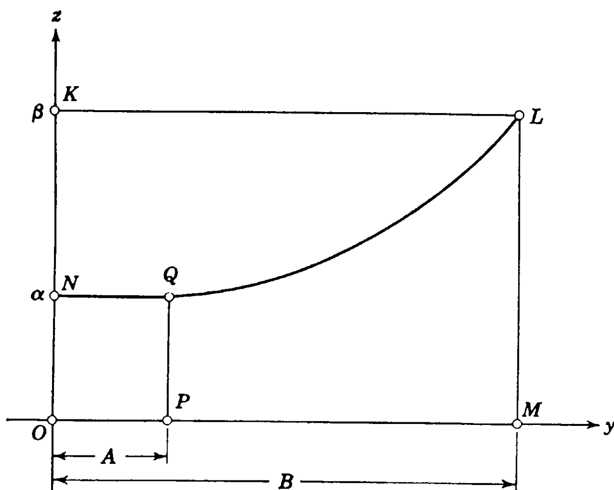


Figura 3.27.

El siguiente ejemplo puede servir como una primera ilustración:

$$\int \log x\,dx = \int \log x \cdot 1\,dx.$$

El integrando se escribe de esta manera para indicar que se hace $f(x) = \log x$ y $g'(x) = 1$, de modo que se tiene $f'(x) = 1/x$ y $g(x) = x$. La fórmula de integración da entonces

$$(55) \qquad \int \log x\,dx = x \log x - \int \frac{x}{x}\,dx = x \log x - x.$$

Esta última expresión es, por lo tanto, la integral indefinida del logaritmo, como puede verificarse de inmediato derivando.

b. Otros ejemplos de integración por partes

Con $f(x) = x$, $g'(x) = e^x$, se tiene $f'(x) = 1$, $g(x) = e^x$ y la integral es entonces

$$\int x\, e^x\, dx = e^x(x - 1). \tag{56}$$

De modo análogo se obtiene

$$\int x \operatorname{sen} x\, dx = -x \cos x + \operatorname{sen} x \tag{57}$$

y

$$\int x \cos x\, dx = x \operatorname{sen} x + \cos x. \tag{58}$$

Para $f(x) = \log x$, $g'(x) = x^a$, se tiene la relación

$$\int x^a \log x\, dx = \frac{x^{a+1}}{a+1}\left(\log x - \frac{1}{a+1}\right). \tag{59}$$

Aquí debe suponerse $a \neq -1$. Para $a = -1$ se obtiene

$$\int \frac{1}{x} \log x\, dx = (\log x)^2 - \int \log x \cdot \frac{dx}{x};$$

y pasando la integral del miembro derecho al izquierdo se tiene [Ver (22), p. 287]

$$\int \frac{1}{x} \log x\, dx = \tfrac{1}{2}(\log x)^2. \tag{60}$$

La integral $\int \operatorname{arc\,sen} x\, dx$ se calcula tomando $f(x) = \operatorname{arc\,sen} x$, $g'(x) = 1$. Por tanto,

$$\int \operatorname{arc\,sen} x\, dx = x \operatorname{arc\,sen} x - \int \frac{x\, dx}{\sqrt{1 - x^2}}.$$

La integración en el miembro derecho puede ser efectuada como en (29), p. 292; se encuentra así

$$\int \operatorname{arc\,sen} x\, dx = x \operatorname{arc\,sen} x + \sqrt{1 - x^2}. \tag{61}$$

De la misma manera se encuentra la integral

$$\int \operatorname{arc\,tan} x\, dx = x \operatorname{arc\,tan} x - \tfrac{1}{2} \log (1 + x^2), \tag{62}$$

y muchas otras de un tipo similar.

Los siguientes ejemplos son de una naturaleza algo diferente. Aquí la repetición de la integración por partes lleva nuevamente a la integral original, para la cual se obtiene así una ecuación.

Procediendo en esta forma se obtiene

$$\int e^{ax} \operatorname{sen} bx\, dx = -\frac{1}{b} e^{ax} \cos bx + \frac{a}{b} \int e^{ax} \cos bx\, dx$$

$$= -\frac{1}{b} e^{ax} \cos bx + \frac{a}{b^2} e^{ax} \operatorname{sen} bx - \frac{a^2}{b^2} \int e^{ax} \operatorname{sen} bx\, dx;$$

Resolviendo esta ecuación para la integral $\int e^{ax} \operatorname{sen} bx\, dx$:

$$\int e^{ax} \operatorname{sen} bx\, dx = \frac{1}{a^2 + b^2} e^{ax}(a \operatorname{sen} bx - b \cos bx). \tag{64}$$

De manera análoga se sigue que

$$\int e^{ax} \cos bx\, dx = \frac{1}{a^2 + b^2} e^{ax}(a \cos bx + b \operatorname{sen} bx). \tag{63}$$

c. Fórmula integral para $f(b) + f(a)$

Como un último ejemplo se deduce una fórmula notable que expresa la suma $f(b) + f(a)$ como una integral definida (en vez de la diferencia $f(b) - f(a)$ dada por la fórmula fundamental). La integración por partes será aplicada introduciendo $1 = g'(x)$, donde $g(x) = x - m$ con una constante m a nuestra disposición. Se tiene entonces para las integrales indefinidas

$$\int f(x)\, dx + \int f'(x)(x - m)\, dx = f(x)(x - m),$$

y para la integral entre a y b

$$\int_a^b f(x)\, dx + \int_a^b f'(x)(x - m)\, dx = f(b)(b - m) - f(a)(a - m).$$

Si con a y b arbitrarios se elige para m el valor medio $m = (a + b)/2$ entre a y b, se obtiene, como el lector puede verificar fácilmente,

$$\frac{b - a}{2}[f(a) + f(b)] = \int_a^b f(x)\, dx + \int_a^b (x - m) f'(x)\, dx.$$

d. Fórmulas recursivas

En muchos casos el integrando no sólo es función de la variable independiente sino que depende también de un índice entero n. Al integrar por partes se obtiene en ocasiones, en vez del valor de la integral, otra expresión similar en la cual el índice n posee un valor menor. Puede llegarse así, después de cierto número de pasos, a una integral que puede ser tratada por medio de la tabla de integrales de la p. 284. Tal proceso es llamado *recursivo.*

Los siguientes ejemplos son ilustraciones: Mediante la repetida integración por partes pueden calcularse las integrales trigonométricas

$$\int \cos^n x\, dx, \qquad \int \operatorname{sen}^n x\, dx, \qquad \int \operatorname{sen}^m x \cos^n x\, dx,$$

siempre que m y n sean enteros positivos. Pues utilizando $f(x) = \cos^{n-1} x$, $g(x) = \operatorname{sen} x$, se encuentra para la primera integral que

$$\int \cos^n x\, dx = \cos^{n-1} x \operatorname{sen} x + (n-1) \int \cos^{n-2} x \operatorname{sen}^2 x\, dx;$$

y el miembro derecho puede ser escrito en la forma

$$\cos^{n-1} x \operatorname{sen} x + (n-1) \int \cos^{n-2} x\, dx - (n-1) \int \cos^n x\, dx.$$

Así se ha obtenido una relación recursiva:

$$(65) \qquad \int \cos^n x\, dx = \frac{1}{n} \cos^{n-1} x \operatorname{sen} x + \frac{n-1}{n} \int \cos^{n-2} x\, dx.$$

Esta fórmula permite disminuir el índice en el integrando paso por paso hasta que finalmente se llega a la integral

$$\int \cos x\, dx = \operatorname{sen} x \qquad \text{o} \qquad \int dx = x,$$

dependiendo de que n sea impar o par. De un modo análogo se obtienen las fórmulas recursivas similares

$$(66) \qquad \int \operatorname{sen}^n x\, dx = -\frac{1}{n} \operatorname{sen}^{n-1} x \cos x + \frac{n-1}{n} \int \operatorname{sen}^{n-2} x\, dx$$

y

$$(67) \qquad \int \operatorname{sen}^m x \cos^n x\, dx = \frac{\operatorname{sen}^{m+1} x \cos^{n-1} x}{m+n} + \frac{n-1}{m+n} \int \operatorname{sen}^m x \cos^{n-2} x\, dx.$$

En particular, se calculan las integrales

$$\int \operatorname{sen}^2 x\, dx = \frac{1}{2} (x - \operatorname{sen} x \cos x)$$

y

$$\int \cos^2 x\, dx = \frac{1}{2} (x + \operatorname{sen} x \cos x),$$

como se ha hecho ya por el método de substitución [Ecs. (42), (43), p. 293].

Casi no es necesario mencionar que las integrales correspondientes para las funciones hiperbólicas pueden ser calculadas de exactamente el mismo modo:

(68) $$\int \operatorname{senh}^2 x\, dx = \frac{1}{2}(-x + \operatorname{senh} x \cosh x),$$

(69) $$\int \cosh^2 x\, dx = \frac{1}{2}(x + \operatorname{senh} x \cosh x).$$

Otras fórmulas recursivas están dadas por las siguientes transformaciones:

(70) $$\int (\log x)^m\, dx = x(\log x)^m - m \int (\log x)^{m-1}\, dx,$$

(71) $$\int x^m e^x\, dx = x^m e^x - m \int x^{m-1} e^x\, dx,$$

(72) $$\int x^m \operatorname{sen} x\, dx = -x^m \cos x + m \int x^{m-1} \cos x\, dx,$$

(73) $$\int x^m \cos x\, dx = x^m \operatorname{sen} x - m \int x^{m-1} \operatorname{sen} x\, dx,$$

(74) $$\int x^a (\log x)^m\, dx = \frac{x^{a+1}(\log x)^m}{a+1} - \frac{m}{a+1}\int x^a (\log x)^{m-1}\, dx \qquad (a \neq -1).$$

e. El producto infinito de Wallis para π

La fórmula recursiva para la integral $\int \operatorname{sen}^n x\, dx$ con $n > 1$ conduce a una expresión fascinante para el número π como "producto infinito". En la fórmula

$$\int \operatorname{sen}^n x\, dx = -\frac{1}{n}\operatorname{sen}^{n-1} x \cos x + \frac{n-1}{n}\int \operatorname{sen}^{n-2} x\, dx$$

se insertan los límites 0 y $\pi/2$, obteniéndose así

(75) $$\int_0^{\pi/2} \operatorname{sen}^n x\, dx = \frac{n-1}{n}\int_0^{\pi/2} \operatorname{sen}^{n-2} x\, dx \qquad \text{para } n > 1.$$

Si repetidamente se aplica la fórmula recursiva, se obtiene, distinguiendo entre los casos $n = 2m$ y $n = 2m + 1$, las integrales

(76) $$\int_0^{\pi/2} \operatorname{sen}^{2m} x\, dx = \frac{2m-1}{2m}\,\frac{2m-3}{2m-2}\cdots\frac{1}{2}\cdot\int_0^{\pi/2} dx,$$

(76a) $$\int_0^{\pi/2} \operatorname{sen}^{2m+1} x\, dx = \frac{2m}{2m+1}\cdot\frac{2m-2}{2m-1}\cdots\frac{2}{3}\cdot\int_0^{\pi/2} \operatorname{sen} x\, dx,$$

por lo que

$$(77)\qquad \int_0^{\pi/2} \operatorname{sen}^{2m} x\, dx = \frac{2m-1}{2m}\cdot\frac{2m-3}{2m-2}\cdots\frac{1}{2}\cdot\frac{\pi}{2},$$

$$(77a)\qquad \int_0^{\pi/2} \operatorname{sen}^{2m+1} x\, dx = \frac{2m}{2m+1}\cdot\frac{2m-2}{2m-1}\cdots\frac{2}{3}.$$

Por división resulta

$$(78)\quad \frac{\pi}{2} = \frac{2\cdot 2}{1\cdot 3}\cdot\frac{4\cdot 4}{3\cdot 5}\cdot\frac{6\cdot 6}{5\cdot 7}\cdots\frac{2m\cdot 2m}{(2m-1)\cdot(2m+1)}\,\frac{\int_0^{\pi/2} \operatorname{sen}^{2m} x\, dx}{\int_0^{\pi/2} \operatorname{sen}^{2m+1} x\, dx}$$

El cociente de las dos integrales en el miembro derecho converge a 1 cuando m crece más allá de toda cota, como se deduce a partir de las siguientes consideraciones. En el intervalo $0 < x < \pi/2$, donde $0 < \operatorname{sen} x < 1$, se tiene

$$0 < \operatorname{sen}^{2m+1} x \leq \operatorname{sen}^{2m} x \leq \operatorname{sen}^{2m-1} x;$$

consecuentemente

$$0 < \int_0^{\pi/2} \operatorname{sen}^{2m+1} x\, dx \leq \int_0^{\pi/2} \operatorname{sen}^{2m} x\, dx \leq \int_0^{\pi/2} \operatorname{sen}^{2m-1} x\, dx.$$

Si se divide aquí cada término por $\int_0^{\pi/2} \operatorname{sen}^{2m+1} x\, dx$ y se observa que, por la fórmula (75), es

$$\frac{\int_0^{\pi/2} \operatorname{sen}^{2m-1} x\, dx}{\int_0^{\pi/2} \operatorname{sen}^{2m+1} x\, dx} = \frac{2m+1}{2m} = 1 + \frac{1}{2m},$$

se tiene que

$$1 \leq \frac{\int_0^{\pi/2} \operatorname{sen}^{2m} x\, dx}{\int_0^{\pi/2} \operatorname{sen}^{2m+1} x\, dx} \leq 1 + \frac{1}{2m},$$

de la cual surge la propiedad antedicha.

Consecuentemente, la relación

$$(79)\qquad \frac{\pi}{2} = \lim_{m\to\infty} \frac{2}{1}\frac{2}{3}\frac{4}{3}\frac{4}{5}\frac{6}{5}\frac{6}{7}\cdots\frac{2m}{2m-1}\frac{2m}{2m+1}$$

se cumple.

Esta fórmula producto (debida a Wallis), con su sencilla ley de formación da una relación notable entre el número π y los enteros.

Producto para $\sqrt{\pi}$

Como una consecuencia sencilla, puede deducirse una expresión igualmente notable para $\sqrt{\pi}$. Si se observa que

$$\lim_{m\to\infty} \frac{2m}{2m+1} = 1,$$

entonces puede escribirse

$$\lim_{m\to\infty} \frac{2^2 \cdot 4^2 \cdots (2m-2)^2}{3^2 \cdot 5^2 \cdots (2m-1)^2} 2m = \frac{\pi}{2},$$

y tomando la raíz cuadrada y multiplicando después el numerador y el denominador por $2 \cdot 4 \cdots (2m-2)$, se encuentra

$$\sqrt{\frac{\pi}{2}} = \lim_{m\to\infty} \frac{2 \cdot 4 \cdots (2m-2)}{3 \cdot 5 \cdots (2m-1)} \sqrt{2m} = \lim_{m\to\infty} \frac{2^2 \cdot 4^2 \cdots (2m-2)^2}{(2m-1)!} \sqrt{2m}$$

$$= \lim_{m\to\infty} \frac{2^2 \cdot 4^2 \cdots (2m)^2}{(2m)!} \frac{\sqrt{2m}}{2m}.$$

$$= \lim \frac{(2^2 \cdot 1^2)(2^2 \cdot 2^2)(2^2 \cdot 3^2) \cdots (2^2 \cdot m^2)}{(2m)!\sqrt{2m}}$$

De aquí se obtiene finalmente

(80) $$\lim_{m\to\infty} \frac{(m!)^2 2^{2m}}{(2m)!\sqrt{m}} = \sqrt{\pi},$$

una forma del producto de Wallis que será utilizada posteriormente (Ver apéndice, capítulo 6).

*3.12 Integración de funciones racionales

Durante los siglos XVII y XVIII los matemáticos estuvieron preocupados por descubrir clases de funciones explícitas elementales que pudieran ser integradas explícitamente. Fueron inventados gran número de medios ingeniosos y al mismo tiempo se creó la base para una comprensión más profunda. Cuando posteriormente se entendió que el lograr la integración de *todas* las funciones explícitas en forma exacta no era una meta alcanzable ni tampoco una meta realmente importante, las tediosas técnicas que se habían desarrollado en conexión con tales problemas fueron gradualmente pasando a segundo plano. Sin embargo, prevaleció un significativo resultado general:

Todas las funciones racionales $R(x)$ de una variable x pueden ser integradas explícitamente en términos de las integrales elementales registradas en la Tabla 3.1.

Este resultado general puede ser obtenido mucho más fácilmente en el contexto de la teoría, más avanzada, de funciones de una variable compleja. No obstante, vale la pena bosquejar una deducción elemental empleando solamente variables reales.

Las funciones racionales son aquéllas de la forma

$$R(x) = \frac{f(x)}{g(x)}, \tag{81}$$

donde $f(x)$ y $g(x)$ son polinomios:

$$f(x) = a_m x^m + a_{m-1}x^{m-1} + \cdots + a_0,$$
$$g(x) = b_n x^n + b_{n-1}x^{n-1} + \cdots + b_0 \quad (b_n \neq 0).$$

Como se recuerda, todo polinomio puede ser integrado de inmediato y su integral es también un polinomio. Basta por ello considerar solamente aquellas funciones racionales para las cuales el denominador $g(x)$ no es una constante. Además, puede suponerse siempre que el grado del numerador es menor que el grado n del denominador. Pues, de lo contrario, dividiendo el polinomio $f(x)$ por el polinomio $g(x)$ se obtiene un residuo de grado menor que n; en otras palabras, puede escribirse $f(x) = q(x)g(x) + r(x)$, donde $q(x)$ y $r(x)$ son también polinomios y $r(x)$ es de grado menor que n. La integración de $f(x)/g(x)$ es reducida entonces a la integración del polinomio $q(x)$ y a la de la fracción "propia" $r(x)/g(x)$. Se observa además que la función $f(x)/g(x)$ puede ser representada como la suma de las funciones $a_\nu x^\nu/g(x)$, de modo que se requiere considerar solamente integrandos de la forma $x^\nu/g(x)$.

a. Los tipos fundamentales

Se procederá por pasos en la integración de las funciones racionales más generales del tipo (81), estudiando primero sólo aquellas funciones con denominador $g(x)$ del tipo particularmente sencillo

$$g(x) = x^n,$$

o bien

$$g(x) = (1 + x^2)^n,$$

donde n es cualquier entero positivo.

A este caso puede reducirse el caso algo más general en el que $g(x) = (\alpha x + \beta)^n$, una potencia entera $(n > 1)$ de una expresión lineal $\alpha x + \beta$ $(\alpha \neq 0)$, o bien $g(x) = (ax^2 + 2bx + c)^n$, una potencia de una expresión cuadrática definida.[1] Si $g(x) = (\alpha x + \beta)^n$, se introduce $\xi = \alpha x + \beta$ como

[1] Una expresión cuadrática $Q(x) = ax^2 + 2bx + c$ se dice que es *definida* si para todos los valores reales de x toma valores que poseen uno y el mismo signo, esto es, si la ecuación $Q(x) = 0$ no posee raíces reales. Para esto es necesario y suficiente que el "discriminante" $ac - b^2$ sea positivo. Esto surge, por supuesto, a partir de la fórmula explícita $(-b \pm \sqrt{b^2 - ac})/a$ para las raíces. Equivalentemente, una expresión cuadrática definida es aquélla que no puede ser factorizada en dos factores lineales *reales*.

una nueva variable. Entonces $d\xi/dx = \alpha$ y $x = (\xi - \beta)/\alpha$ es también una función lineal de ξ. Cada numerador $f(x)$ se vuelve un polinomio $\phi(\xi)$ del mismo grado, y, consecuentemente,

$$\int \frac{f(x)}{(\alpha x + \beta)^n}\, dx = \frac{1}{\alpha}\int \frac{\phi(\xi)}{\xi^n}\, d\xi.$$

En el segundo caso se escribe

$$ax^2 + 2bx + c = \frac{1}{a}(ax + b)^2 + \frac{d^2}{a} \qquad (d^2 = ac - b^2, d > 0);$$

y como se ha supuesto que la expresión es cuadrática y definida, $ac - b^2$ debe ser positivo y $a \neq 0$. Introduciendo la nueva variable

$$\xi = \frac{ax + b}{d}$$

se obtiene una integral con denominador $[(d^2/a)(1 + \xi^2)]^n$, en el integrando.

Por lo tanto, para integrar funciones racionales cuyos denominadores son potencias de una expresión lineal o de expresiones cuadráticas definidas es suficiente con poder integrar los siguientes tipos de funciones:

$$\frac{1}{x^n}, \quad \frac{x^{2\nu}}{(x^2 + 1)^n}, \quad \frac{x^{2\nu+1}}{(x^2 + 1)^n}.$$

Se verá, en efecto, que aún estos tipos no requieren ser tratados en general, pues se puede reducir la integración de cada función racional a la integración de las formas muy especiales obtenidas de estas tres funciones tomando $\nu = 0$. De acuerdo con esto se considerará ahora la integración de las tres expresiones

$$\frac{1}{x^n}, \quad \frac{1}{(x^2 + 1)^n}, \quad \frac{x}{(x^2 + 1)^n}.$$

b. Integración de los tipos fundamentales

La integración del primer tipo de función, $1/x^n$, conduce inmediatamente a la expresión $\log|x|$ si $n = 1$, y a la expresión $-1/(n - 1)x^{n-1}$ si $n > 1$, de modo que en ambos casos la integral es también una función elemental. Las funciones del tercer tipo pueden ser integradas rápidamente introduciendo la nueva variable $\xi = x^2 + 1$, de lo cual se obtiene $2x\, dx = d\xi$ y

$$\int \frac{x}{(x^2 + 1)^n}\, dx = \frac{1}{2}\int \frac{d\xi}{\xi^n} = \begin{cases} \frac{1}{2}\log(x^2 + 1) & \text{si } n = 1, \\ -\dfrac{1}{2(n - 1)(x^2 + 1)^{n-1}} & \text{si } n > 1. \end{cases}$$

Finalmente, para calcular la integral

$$I_n = \int \frac{dx}{(x^2+1)^n},$$

donde n tiene cualquier valor entero mayor que uno, se hace uso del método recursivo: Si se pone

$$\frac{1}{(x^2+1)^n} = \frac{1}{(x^2+1)^{n-1}} - \frac{x^2}{(x^2+1)^n},$$

de modo que

$$\int \frac{dx}{(x^2+1)^n} = \int \frac{dx}{(x^2+1)^{n-1}} - \int \frac{x^2\,dx}{(x^2+1)^n},$$

puede transformarse el miembro derecho mediante integración por partes, utilizando la fórmula (54) de la p. 295 con

$$f(x) = x, \qquad g'(x) = \frac{x}{(x^2+1)^n}.$$

Entonces, tal como se ha encontrado,

$$g(x) = -\frac{1}{2}\,\frac{1}{(n-1)(x^2+1)^{n-1}},$$

y, consecuentemente, se obtiene

$$I_n = \int \frac{dx}{(x^2+1)^n} = \frac{x}{2(n-1)(x^2+1)^{n-1}} + \frac{2n-3}{2(n-1)}\int \frac{dx}{(x^2+1)^{n-1}}.$$

El cálculo de la integral I_n es reducido así al de la integral I_{n-1}. Si $n-1>1$, se aplica el mismo proceso a esta última integral y se continúa hasta que finalmente se llega a la expresión

$$\int \frac{dx}{x^2+1} = \operatorname{arc\,tan} x.$$

Se ve así que la integral[1] I_n puede ser expresada explícitamente en términos de funciones racionales y de la función arc tan x.

Por otra parte, nótese que pudo haberse integrado la función $1/(x^2+1)^n$ también en forma directa, utilizando la substitución $x = \tan t$; se habría obtenido entonces $dx = \sec^2 t\,dt$ y $1/(1+x^2) = \cos^2 t$, de modo que

$$\int \frac{dx}{(x^2+1)^n} = \int \cos^{2n-2} t\,dt,$$

y se ha aprendido ya [Ec. (65), p. 299] cómo evaluar esta integral.

---

[1] La integral de la función $1/(x^2-1)^n$ puede ser calculada de la misma manera; por el correspondiente método de recurrencia se reduce ésta a la integral

$$\int \frac{dx}{1-x^2} = \operatorname{ar\,tanh} x \qquad (\text{o bien } \operatorname{ar\,coth} x).$$

c. Fracciones parciales

Se está ahora en condiciones de integrar las funciones racionales más generales. Se hace uso del hecho de que cualquiera de tales funciones puede ser representada como la suma de las así llamadas *fracciones parciales,* esto es, como la suma de un polinomio y un número finito de funciones racionales cada una de las cuales tiene, ya sea una potencia de una expresión lineal como denominador y una constante definida como numerador, o bien una potencia de una expresión cuadrática definida como denominador y una función lineal como numerador. Si el grado del numerador $f(x)$ es menor que el del denominador $g(x)$, el polinomio no aparece. Se sabe ya cómo integrar cada fracción parcial. Pues, de acuerdo con lo visto en la p. 304, el denominador puede ser reducido a una de las formas especiales x^n y $(x^2+1)^n$, y la fracción es entonces una combinación de los tipos fundamentales integrados en la p. 304.

No se dará la demostración general de la posibilidad de esta resolución en fracciones parciales. Simplemente se tratará de hacer inteligible para el lector el contenido del teorema y mostrar mediante ejemplos cómo la resolución en fracciones parciales puede ser llevada a cabo en casos típicos. En la práctica se tratan solamente funciones comparativamente sencillas, pues de otra forma los cálculos se tornan demasiado complicados.

Como se sabe del álgebra elemental, todo polinomio real $g(x)$ puede ser escrito en la forma [1]

$$g(x) = a(x-\alpha_1)^{l_1}(x-\alpha_2)^{l_2}\cdots$$
$$\cdots(x^2+2b_1x+c_1)^{r_1}(x^2+2b_2x+c_2)^{r_2}\cdots.$$

Aquí los números distintos $\alpha_1, \alpha_2, \ldots$ son las raíces reales de la ecuación $g(x)=0$, y los enteros positivos $l_1, l_2, \ldots$ indican la multiplicidad de estas raíces; los factores $x^2+2b_\nu x+c_\nu$ indican expresiones cuadráticas definidas, de las cuales no hay dos iguales, con raíces complejas conjugadas, y los enteros postitivos $r_1, r_2, \ldots$ dan la multiplicidad de estas raíces.

Se supone que el denominador está dado en esta forma, o bien que ha sido llevado a esta forma calculando las raíces reales e imaginarias. Se supone además que el numerador $f(x)$ es de menor grado que el denominador (Ver p. 303). Entonces el teorema sobre resolución en fracciones parciales puede establecerse como sigue: *Para cada factor* $(x-\alpha)^l$, *donde* α *es cualquiera de las raíces reales de multiplicidad* l, *se puede determinar una expresión de la forma*

[1] La *demostración* real de este teorema, llamado *teorema fundamental del álgebra,* no pertenece al álgebra. Se la logra más fácilmente por métodos que pertenecen a la teoría de funciones de una *variable compleja.*

$$\frac{A_1}{x-\alpha}+\frac{A_2}{(x-\alpha)^2}+\cdots+\frac{A_l}{(x-\alpha)^l},$$

y para cada factor cuadrático $Q(x) = x^2 + 2bx + c$ *en el producto, que está elevado a la potencia r, se puede determinar una expresión de la forma*

$$\frac{B_1+C_1x}{Q}+\frac{B_2+C_2x}{Q^2}+\cdots+\frac{B_r+C_rx}{Q^r},$$

de tal manera que la función $f(x)/g(x)$ *es la suma de todas estas expresiones* (A_ν, B_ν, C_ν, *son constantes*). En otras palabras, el cociente $f(x)/g(x)$ puede ser representado como una suma de fracciones cada una de las cuales pertenece a uno de los tipos integrados arriba.[1]

En casos particulares la descomposición en fracciones parciales puede efectuarse fácilmente por inspección. Si, por ejemplo, $g(x) = x^2 - 1$, se ve de inmediato que

$$\frac{1}{x^2-1}=\frac{1}{2}\,\frac{1}{(x-1)}-\frac{1}{2}\,\frac{1}{(x+1)},$$

de modo que

$$\int\frac{dx}{x^2-1}=\tfrac{1}{2}\log\left|\frac{x-1}{x+1}\right|.$$

Más generalmente, si $g(x) = (x-\alpha)(x-\beta)$, esto es, si $g(x)$ es una expresión cuadrática no definida con dos ceros reales α y β, se tiene

$$\frac{1}{(x-\alpha)(x-\beta)}=\frac{1}{(\alpha-\beta)}\,\frac{1}{(x-\alpha)}-\frac{1}{(\alpha-\beta)}\,\frac{1}{(x-\beta)},$$

[1] Bosquejamos brevemente un método mediante el cual puede probarse la posibilidad de esta descomposición en fracciones parciales sin usar la teoría de funciones de variables complejas, suponiendo que $g(x)$ puede factorizarse completamente en factores lineales. Si $g(x) = (x-\alpha)^k h(x)$ y $h(\alpha) \neq 0$, entonces el numerador en el lado derecho de la ecuación

$$\frac{f(x)}{g(x)}-\frac{f(\alpha)}{h(\alpha)(x-\alpha)^k}=\frac{1}{h(\alpha)}\,\frac{f(x)h(\alpha)-f(\alpha)h(x)}{(x-\alpha)^k h(x)}$$

obviamente se anula para $x = \alpha$; por consiguiente, es de la forma $h(\alpha)(x-\alpha)^m f_1(x)$, donde $f_1(x)$ es también un polinomio, el entero m es ≥ 1 y $f_1(\alpha) \neq 0$. Escribiendo $f(\alpha)/h(\alpha) = \beta$, la anterior da

$$\frac{f(x)}{g(x)}-\frac{\beta}{(x-\alpha)^k}=\frac{f_1(x)}{(x-\alpha)^{k-m}h(x)}.$$

Continuando este proceso, podemos hacer que disminuya paso a paso el grado de la potencia de $(x-\alpha)$ que aparece en el denominador, hasta que finalmente no aparezca tal factor. Repetimos el proceso con la fracción restante considerando alguna otra raíz de $g(x)$, y hacemos lo mismo tantas veces como factores distintos tenga $g(x)$. Haciendo esto no sólo para las raíces reales sino también para las complejas, y combinando fracciones complejas conjugadas, finalmente se llegará a la descomposición completa en fracciones parciales.

de modo que

$$\int \frac{dx}{(x-\alpha)(x-\beta)} = \frac{1}{\alpha-\beta} \log \left| \frac{x-\alpha}{x-\beta} \right|$$

d. Ejemplos de resolución en fracciones parciales. El método de los coeficientes indeterminados

Si $g(x) = (x-\alpha_1)(x-\alpha_2)\cdots(x-\alpha_n)$, donde $\alpha_i \neq \alpha_k$ si $i \neq k$, esto es, si la ecuación $g(x) = 0$ posee solamente raíces reales simples, y si $f(x)$ es cualquier polinomio de grado $<n$, la expresión en términos de fracciones parciales tiene la sencilla forma

$$\frac{f(x)}{g(x)} = \frac{a_1}{x-\alpha_1} + \frac{a_2}{x-\alpha_2} + \cdots + \frac{a_n}{x-\alpha_n}.$$

Se obtienen expresiones explícitas para los coeficientes $a_1, a_2, \ldots$ si se multiplican ambos lados de esta ecuación por $(x-\alpha_1)$, se cancela el factor común $(x-\alpha_1)$ en el numerador y el denominador de la izquierda y en el primer término de la derecha, y luego se hace $x = \alpha_1$. Esto da

$$a_1 = \frac{f(\alpha_1)}{(\alpha_1-\alpha_2)(\alpha_1-\alpha_3)\cdots(\alpha_1-\alpha_n)}.$$

El lector observará a partir de la regla para la derivada de un producto que el denominador a la derecha es $g'(\alpha_1)$, esto es, la derivada de la función $g(x)$ en el punto $x = \alpha_1$. Fórmulas análogas para $a_2, a_3, \ldots$, obtenidas de esta manera, conducen a la expresión explícita en fracciones parciales

$$\frac{f(x)}{g(x)} = \frac{f(\alpha_1)}{g'(\alpha_1)(x-\alpha_1)} + \frac{f(\alpha_2)}{g'(\alpha_2)(x-\alpha_2)} + \cdots + \frac{f(\alpha_n)}{g'(\alpha_n)(x-\alpha_n)}.$$

Como un ejemplo típico de un denominador $g(x)$ con raíces múltiples se considera la función $1/[x^2(x-1)]$. Esta tiene por representación

$$\frac{1}{x^2(x-1)} = \frac{a}{x-1} + \frac{b}{x} + \frac{c}{x^2},$$

en concordancia con la p. 307. Si se multiplican ambos miembros de esta ecuación por $x^2(x-1)$, se obtiene la ecuación

$$1 = (a+b)x^2 - (b-c)x - c,$$

válida para todos los valores de x, a partir de la cual han de determinarse los coeficientes a, b, c. Esta condición no puede cumplirse a menos que todos los coeficientes del polinomio $(a+b)x^2 - (b-c)x - c - 1$ sean cero; esto es, se debe tener $a+b = b-c = c+1 = 0$, o bien $c = -1$, $b = -1$, $a = 1$. Se obtiene así la resolución

$$\frac{1}{x^2(x-1)} = \frac{1}{x-1} - \frac{1}{x} - \frac{1}{x^2},$$

y, consecuentemente,

$$\int \frac{dx}{x^2(x-1)} = \log|x-1| - \log|x| + \frac{1}{x}.$$

A continuación se descompone la función $1/[x(x^2+1)]$, cuyo denominador posee ceros complejos, de acuerdo con la ecuación

$$\int \frac{1}{x(x^2+1)} = \frac{a}{x} + \frac{bx+c}{x^2+1}.$$

Para los coeficientes se obtiene $a + b = c = a - 1 = 0$, de modo que

$$\frac{1}{x(x^2+1)} = \frac{1}{x} - \frac{x}{x^2+1},$$

y, por lo tanto,

$$\int \frac{dx}{x(x^2+1)} = \log|x| - \tfrac{1}{2}\log(x^2+1).$$

Como un tercer ejemplo se considera la función $1/(x^4+1)$, cuya integración era un serio problema aun en los tiempos de Leibnitz. Se puede representar el denominador como el producto de dos factores cuadráticos:[1]

$$x^4 + 1 = (x^2+1)^2 - 2x^2 = (x^2 + 1 + \sqrt{2}x)(x^2 + 1 - \sqrt{2}x).$$

Se sabe, por lo tanto, que la resolución en fracciones parciales tendrá la forma

$$\frac{1}{x^4+1} = \frac{ax+b}{x^2+\sqrt{2}x+1} + \frac{cx+d}{x^2-\sqrt{2}x+1}.$$

Para determinar los coeficientes a, b, c, d, se utiliza la ecuación

$$(a+c)x^3 + (b + d - a\sqrt{2} + c\sqrt{2})x^2 + (a + c - b\sqrt{2} + d\sqrt{2})x + (b + d - 1) = 0,$$

la cual se satisface para los valores

$$a = \frac{1}{2\sqrt{2}}, \qquad b = \tfrac{1}{2}, \qquad c = -\frac{1}{2\sqrt{2}}, \qquad d = \tfrac{1}{2}.$$

Se tiene entonces

$$\frac{1}{x^4+1} = \frac{1}{2\sqrt{2}}\cdot\frac{x+\sqrt{2}}{x^2+\sqrt{2}x+1} - \frac{1}{2\sqrt{2}}\cdot\frac{x-\sqrt{2}}{x^2-\sqrt{2}x+1},$$

[1] La factorización de $x^4 + 1$ en factores cuadráticos reales corresponde a la factorización en factores lineales complejos conjugados

$$x^4 + 1 = [(x-\varepsilon)(x-\varepsilon^{-1})][(x-\varepsilon^3)(x-\varepsilon^{-3})],$$

donde

$$\varepsilon = \cos\frac{\pi}{4} + i\,\mathrm{sen}\,\frac{\pi}{4} = \frac{1}{2}\sqrt{2}(1+i)$$

es una de las ocho raíces de $+1$, y una cuarta raíz de -1 (véase pp. 128-129).

y, aplicando el método dado en la p. 304, se obtiene

$$\int \frac{dx}{x^4+1} = \frac{1}{4\sqrt{2}}\log|x^2+\sqrt{2}x+1| - \frac{1}{4\sqrt{2}}\log|x^2-\sqrt{2}x+1|$$

$$+\frac{1}{2\sqrt{2}}\text{arc tan }(\sqrt{2}x+1)+\frac{1}{2\sqrt{2}}\text{arc tan }(\sqrt{2}x-1),$$

lo cual puede ser verificado fácilmente por derivación.

Los ejemplos anteriores ilustran un método general de integrar una función racional $f(x)/g(x)$. Se divide primero y se reduce el problema al caso en que el grado de f es menor que el de g. Se descompone $g(x)$ en factores lineales cuadráticos definidos, agrupando el producto en potencias de tales factores. Se escribe la representación en fracciones parciales *apropiadas* de f/g, con coeficientes *indeterminados* a, b, c, ... Multiplicando todo por $g(x)$ y *comparando coeficientes* de potencias iguales en la identidad polinomial resultante se obtiene un sistema de ecuaciones lineales para los coeficientes desconocidos, que justamente debe ser adecuado para determinar estos coeficientes si realmente se tiene la forma correcta para la descomposición en fracciones parciales. Se está entonces en posibilidad de integrar cualquiera de las fracciones parciales resultantes mediante las reglas discutidas anteriormente.

3.13 Integración de algunas otras clases de funciones

a. Comentarios preliminares sobre la representación[1] racional de la circunferencia y de la hipérbola

La integración de algunas otras clases generales de funciones puede ser reducida a la integración de funciones racionales. Se estará en condiciones de entender mejor esta reducción estableciendo primero ciertas propiedades elementales de las funciones trigonométricas e hiperbólicas. Si se hace $t = \tan(x/2)$, la trigonometría elemental conduce a las fórmulas sencillas

$$\operatorname{sen} x = \frac{2t}{1+t^2}, \qquad \cos x = \frac{1-t^2}{1+t^2};$$

en efecto, a partir de

$$\frac{1}{1+t^2} = \cos^2\frac{x}{2} \qquad \text{y} \qquad \frac{t^2}{1+t^2} = \operatorname{sen}^2\frac{x}{2},$$

y de las fórmulas elementales

$$\operatorname{sen} x = 2\cos^2\frac{x}{2}\tan\frac{x}{2} \qquad \text{y} \qquad \cos x = \cos^2\frac{x}{2} - \operatorname{sen}^2\frac{x}{2},$$

se obtienen esas ecuaciones. Ellas muestran que $\operatorname{sen} x$ *y* $\cos x$ *pueden ser ambas expresadas racionalmente en términos de la cantidad* $t = \tan x/2$. Derivando se tiene

$$\frac{dt}{dx} = \frac{1}{2\cos^2 x/2} = \frac{1+t^2}{2},$$

[1] Llamada a veces "uniformización".

de modo que

$$\frac{dx}{dt}=\frac{2}{1+t^2}; \tag{82}$$

por lo que la derivada dx/dt es también una expresión racional en t.

*La representación geométrica de nuestras fórmulas y su significado geométrico se da en la Fig. 3.28. Aquí se muestra el círculo $u^2 + v^2 = 1$ en un plano u, v. Si x denota el ángulo TOP en la figura, entonces $u = \cos x$ y $v = \operatorname{sen} x$. El ángulo OSP, con su vértice en el punto $u = -1$, $v = 0$, es igual a $x/2$, por un teorema de la geometría elemental, y puede extraerse el significado geométrico del parámetro t de la figura; $t = \tan \frac{1}{2}x = OR$ donde R es la "proyección" sobre el eje v del punto

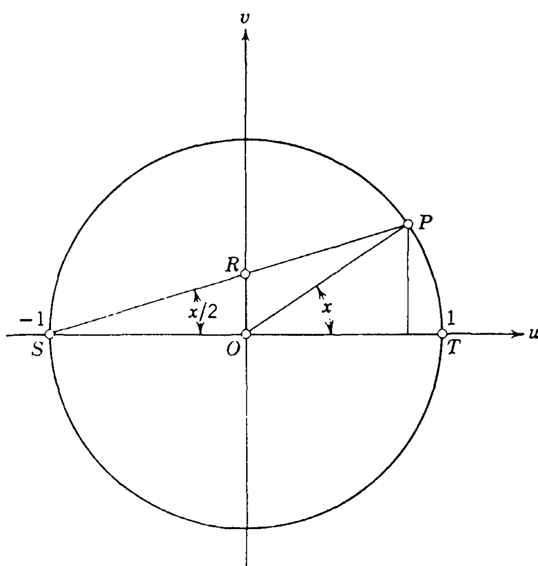


Figura 3.28 Representación paramétrica de las funciones trigonométricas.

P del círculo desde S. Si el punto P parte desde S y describe una vez el círculo en la dirección positiva, esto es, si x recorre el intervalo desde $-\pi$ hasta $+\pi$, la cantidad t recorrerá todo el rango de valores desde $-\infty$ hasta $+\infty$ una vez exactamente. (Nótese que el punto S mismo corresponde a $t = \pm\infty$.) Se tiene aquí una representación del punto general (u, v) del círculo $u^2 + v^2 = 1$ en términos de las funciones racionales $u = (1 - t^2)/(1 + t^2)$, y $v = 2t/(1 + t^2)$ del parámetro t. Estas fórmulas definen entonces una transformación racional de la recta t sobre un círculo del plano u, v (la cual es la análoga bidimensional de la proyección estereográfica sobre una esfera, mencionada en la p. 45). En la base de esta representación racional del círculo se encuentra, evidentemente, la identidad

$$(t^2 - 1)^2 + (2t)^2 = (t^2 + 1)^2.$$

Es por demás curioso que esta fórmula sea de interés también en la teoría de números, puesto que genera para cada entero t *enteros pitagóricos* $a = t^2 - 1$, $b = 2t$, y $c = t^2 + 1$ que satisfacen la identidad $a^2 + b^2 = c^2$; esto es, determina un trián-

gulo rectángulo con lados conmensurables. Así $t = 2$ da lugar a la terna bien conocida $a = 3$, $b = 4$, $c = 5$; para $t = 4$ se obtiene $a = 15$, $b = 8$, $c = 17$, etc. Es notable, y, por supuesto, no accidental, que la misma identidad algebraica es de significación en contextos tan diversos como la integración en forma exacta, la geometría y la teoría de números. Ligar campos diferentes de tal manera es una tendencia típica en las matemáticas modernas; aunque cuando el ejemplo en cuestión se remonta a la antigüedad.

Análogamente pueden expresarse las funciones hiperbólicas

$$\cosh x = \tfrac{1}{2}(e^x + e^{-x})$$

y $\operatorname{senh} x = \tfrac{1}{2}(e^x - e^{-x})$ como funciones racionales de una tercera cantidad. El camino más obvio es hacer $e^x = \tau$, de modo que se tiene

$$\cosh x = \frac{1}{2}\left(\tau + \frac{1}{\tau}\right), \qquad \operatorname{senh} x = \frac{1}{2}\left(\tau - \frac{1}{\tau}\right),$$

que son expresiones racionales para $\operatorname{senh} x$ y $\cosh x$. Aquí, nuevamente, $dx/d\tau = 1/\tau$ es racional en τ. Sin embargo, se obtiene una analogía más cercana con las funciones trigonométricas introduciendo la cantidad $t = \tanh(x/2) = (\tau - 1)/(\tau + 1)$; y se llega entonces a las fórmulas

$$\cosh x = \frac{1 + t^2}{1 - t^2}, \qquad \operatorname{senh} x = \frac{2t}{1 - t^2}.$$

Derivando $t = \tanh(x/2)$ se obtiene, como en la Ec. (82) de la p. 311, la expresión racional

$$\frac{dx}{dt} = \frac{2}{1 - t^2} \tag{83}$$

para la derivada dx/dt. Aquí, nuevamente, la cantidad t posee un significado geométrico similar al que tiene para las funciones trigonométricas, como se ve de inmediato de la Fig. 3.29.

Se tiene aquí una representación racional de la hipérbola $u^2 - v^2 = 1$ en el plano u, v, por medio de las ecuaciones $u = (1 + t^2)/(1 - t^2)$, $v = 2t/(1 - t^2)$. Los puntos en la rama derecha de la curva son de la forma $u = \cosh x$, $v = \operatorname{senh} x$, y corresponden a valores de t con $|t| < 1$. La otra rama se obtiene para $|t| > 1$.

Se prosigue ahora con los problemas de integración.

*b. Integración de $R(\cos x, \operatorname{sen} x)$

Denótese por $R(\cos x, \operatorname{sen} x)$ una expresión que es racional en las dos funciones $\operatorname{sen} x$ y $\cos x$, esto es, una expresión que se forma racionalmente con estas dos funciones y algunas constantes, tal como

$$\frac{3 \operatorname{sen}^2 x + \cos x}{3 \cos^2 x + \operatorname{sen} x}$$

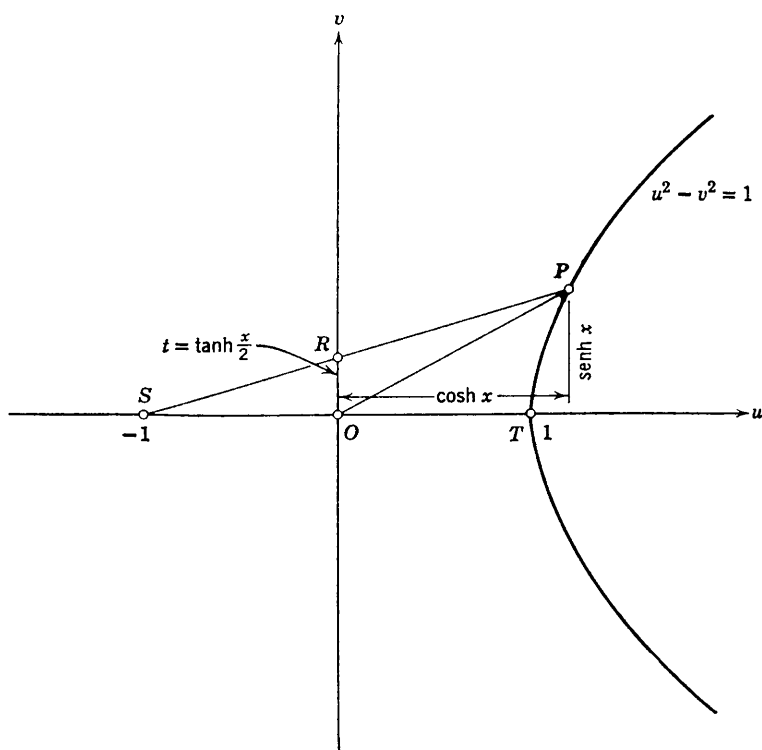


Figura 3.29 Representación paramétrica de las funciones hiperbólicas.

Si se aplica la substitución $t = \tan x/2$, la integral

$$\int R(\cos x, \operatorname{sen} x)\, dx$$

es transformada en la integral

$$\int R\left(\frac{1-t^2}{1+t^2}\ \frac{2t}{1+t^2}\right)\frac{2}{1+t^2}\, dt,$$

y bajo el signo de integral se tiene ahora una función racional de t. Así se ha obtenido, en principio, la integral de nuestra expresión, puesto que puede efectuarse ahora la integración por los métodos de la sección anterior.

c. Integración de $R(\cosh x, \operatorname{senh} x)$

Del mismo modo, si $R(\cosh x, \operatorname{senh} x)$ es una expresión racional en términos de las funciones hiperbólicas $\cosh x$ y $\operatorname{senh} x$, puede efectuarse su integración por medio de la substitución $t = \tanh x/2$. Recordando la Ec. (83), se tiene

$$\int R(\cosh x, \operatorname{senh} x)\, dx = \int R\left(\frac{1+t^2}{1-t^2}, \frac{2t}{1-t^2}\right)\frac{2}{1-t^2}\, dt.$$

(De acuerdo con un comentario anterior pudo haberse introducido también $\tau = e^x$ como una nueva variable y expresado $\cosh x$ y $\operatorname{senh} x$ en términos de τ). La integración queda nuevamente reducida a la de una función racional.

*d. Integración de $R(x, \sqrt{1-x^2})$

La integral $\int R(x, \sqrt{1-x^2})\, dx$ puede ser reducida a la del tipo tratado en la sección 3.13*b* utilizando la substitución

$$x = \cos u, \qquad \sqrt{1-x^2} = \operatorname{sen} u, \qquad dx = -\operatorname{sen} u\, du.$$

A partir de este paso la transformación $t = \tan u/2$ conduce a la integración de una función racional. Obsérvese que pudo haberse efectuado la reducción en un paso, en vez de dos, utilizando la substitución

$$t = \sqrt{\frac{1-x}{1+x}}; \qquad x = \frac{1-t^2}{1+t^2};$$

$$\sqrt{1-x^2} = \frac{2t}{1+t^2}; \qquad \frac{dx}{dt} = \frac{-4t}{(1+t^2)^2};$$

esto es, pudo haberse introducido $t = \tan u/2$ directamente como la nueva variable y con ello obtenido un integrando racional.

*e. Integración de $R(x, \sqrt{x^2-1})$

La integral $\int R(x, \sqrt{x^2-1})\, dx$ es transformada mediante la substitución $x = \cosh u$ en una del tipo tratado en la sección 3.13*c*. Aquí también puede llegarse a la meta directamente introduciendo

$$t = \sqrt{\frac{x-1}{x+1}} = \tanh\frac{u}{2}.$$

*f. Integración de $R(x, \sqrt{x^2+1})$

La integral $\int R(x, \sqrt{(x^2+1)})\, dx$ es reducida mediante la transformación $x = \operatorname{senh} u$ a una del tipo considerado en la sección 3.13*c* (pp. 313-314) y puede, por ello, ser integrada en términos de funciones elementales. En vez de la reducción posterior a la integral de una función racional por medio de la substitución $e^u = \tau$, o de $\tanh u/2 = t$, podría llegarse a la integral de una función racional en un solo paso por cualquiera de las substituciones

$$x = \frac{\xi^2 - \beta}{\alpha} \quad \text{y} \quad \frac{dx}{d\xi} = \frac{2\xi}{\alpha};$$

*g. Integración de $R(x, \sqrt{ax^2 + 2bx + c})$

La integral $\int R(x, \sqrt{ax^2 + 2bx + c})\, dx$ de una expresión que es racional en términos de x y de la raíz cuadrada de un polinomio arbitrario de segundo grado en x puede ser reducida inmediatamente a uno de los tipos tratados. Se escribe (Ver p. 304)

$$ax^2 + 2bx + c = \frac{1}{a}(ax + b)^2 + \frac{ac - b^2}{a}.$$

Si $ac - b^2 > 0$, se introduce una nueva variable ξ por medio de la transformación $\xi = (ax + b)/\sqrt{ac - b^2}$, después de lo cual el radical adquiere la forma $\sqrt{(ac - b^2)(\xi^2 + 1)/a}$. Por lo tanto, la integral cuando está expresada en términos de ξ es del tipo de la sección 3.13*f*. La constante a debe ser aquí positiva de modo que la raíz cuadrada pueda tener valores reales.

Si $ac - b^2 = 0$ y $a > 0$, entonces por medio de la fórmula

$$\sqrt{ax^2 + 2bx + c} = \sqrt{a}\left(x + \frac{b}{a}\right)$$

se ve que el integrando era racional desde un principio.

Si, finalmente, $ac - b^2 < 0$, se hace $\xi = (ax + b)/\sqrt{b^2 - ac}$ y se obtiene para el radical la expresión $\sqrt{(ac - b^2)(\xi^2 - 1)/a}$. Si a es positiva, la integral es así reducida al tipo de la sección 3.13*e*; si, por otra parte, a es negativa el radical se escribe en la forma

$$\sqrt{(b^2 - ac)(1 - \xi^2)/(-a)},$$

y se ve que la integral es así reducida al tipo de la sección 3.13*d*.

*h. Otros ejemplos de reducción a integrales de funciones racionales

De otros tipos de funciones que pueden ser integradas mediante reducción a funciones racionales, serán mencionados solamente dos: 1) expresiones racionales que involucran dos raíces cuadradas diferentes de expresiones lineales, $R(x, \sqrt{ax + b}, \sqrt{\alpha x + \beta})$; 2) expresiones de la forma $R(x, \sqrt[n]{(ax + b)/(\alpha x + \beta)})$, donde a, b, α, β son constantes. En las del primer tipo se introduce la nueva variable $\xi = \sqrt{\alpha x + \beta}$, de modo que $\alpha x + \beta = \xi^2$, y, consecuentemente,

$$x = \frac{\xi^2 - \beta}{\alpha} \quad \text{y} \quad \frac{dx}{d\xi} = \frac{2\xi}{\alpha};$$

entonces

$$\int R(x, \sqrt{ax+b}, \sqrt{\alpha x+\beta})\, dx$$

$$= \int R\left\{\frac{\xi^2-\beta}{\alpha}, \sqrt{\frac{1}{\alpha}[a\xi^2-(a\beta-b\alpha)]}, \xi\right\}\frac{2\xi}{\alpha}\, d\xi,$$

que es del tipo discutido en la sección 3.13*g*.

Si en las del segundo tipo se introduce la nueva variable

$$\xi = \sqrt[n]{\frac{ax+b}{\alpha x+\beta}},$$

se tiene

$$\xi^n = \frac{ax+b}{\alpha x+\beta}, \qquad x = \frac{-\beta\xi^n+b}{\alpha\beta^n-a}, \qquad \frac{dx}{d\xi} = \frac{a\beta-b\alpha}{(\alpha\xi^n-a)^2}\, n\xi^{n-1},$$

e inmediatamente se llega a la fórmula

$$\int R\left(x, \sqrt[n]{\frac{ax+b}{\alpha x+\beta}}\right) dx = \int R\left(\frac{-\beta\xi^n+b}{\alpha\xi^n-a}, \xi\right)\frac{a\beta-b\alpha}{(a\xi^n-a)^2}\, n\xi^{n-1}\, d\xi,$$

que es la integral de una función racional.

i. Comentarios sobre los ejemplos

Las discusiones anteriores son principalmente de interés teórico. En expresiones complicadas los cálculos reales podrían ser demasiado intrincados. Es, por lo tanto, conveniente aprovechar cuando sea posible la forma especial del integrando para simplificar el trabajo. Por ejemplo, para integrar $1/(a^2 \operatorname{sen}^2 x + b^2 \cos^2 x)$ es mejor utilizar la substitución $t = \tan x$ en vez de la dada en las pp. 313-314; pues $\operatorname{sen}^2 x$ y $\cos^2 x$ pueden ser expresados racionalmente en términos de $\tan x$, y es entonces innecesario retroceder a $t = \tan x/2$. Lo mismo es cierto para toda expresión formada racionalmente a partir de [1] $\operatorname{sen}^2 x$, $\cos^2 x$ y $\operatorname{sen} x \cos x$. Además, para el cálculo de muchas integrales es preferible una forma trigonométrica a una racional, siempre que la forma trigonométrica pueda ser evaluada por algún método sencillo de recurrencia. Por ejemplo, aunque el integrando en $\int x^n(\sqrt{1-x^2})^m\, dx$ puede ser reducido a una forma racional, es mejor escribir $x = \operatorname{sen} u$ y llevarla a la forma $\int \operatorname{sen}^n u \cos^{m+1} u\, du$, puesto que ésta puede ser tratada fácilmente por el método de recurrencia de la p. 299 (o bien, utilizando los teoremas de adición para reducir las potencias del seno y coseno a senos y cosenos de ángulos múltiples).

---

[1] Pues $\operatorname{sen} x \cos x = \tan x \cos^2 x$ puede, por supuesto, ser expresado racionalmente en términos de $\tan x$.

Para la evaluación de la integral

$$\int \frac{dx}{a \cos x + b \operatorname{sen} x} \qquad (a^2 + b^2 > 0),$$

en vez de recurrir a la teoría general se escribe

$$A = \sqrt{a^2 + b^2}, \qquad \operatorname{sen} \theta = \frac{a}{A}, \qquad \cos \theta = \frac{b}{A}.$$

La integral toma entonces la forma

$$\frac{1}{A} \int \frac{dx}{\operatorname{sen} (x + \theta)},$$

e introduciendo la nueva variable $x + \theta$ se encuentra [(cfr. Ec. (40), p. 293)] que el valor de la integral es

$$\frac{1}{A} \log \left| \tan \frac{x + \theta}{2} \right|.$$

Parte C Otros pasos en la teoría del cálculo integral

3.14 Integrales de funciones elementales

a. Definición de funciones por medio de integrales. Integrales y funciones elípticas

Con los ejemplos que se han dado de los tipos de funciones que pueden ser integradas mediante la reducción a funciones racionales se ha terminado prácticamente la lista de funciones que son integrables en términos de funciones elementales. Intentos de expresar integrales indefinidas tales como (para $n > 2$)

$$\int \frac{dx}{\sqrt{a_0 + a_1 x + \cdots + a_n x^n}},$$

$$\int \sqrt{a_0 + a_1 x + \cdots + a_n x^n}\, dx,$$

o bien

$$\int \frac{e^x}{x}\, dx$$

en términos de funciones elementales han fallado. En el siglo XIX finalmente se probó que es realmente *imposible* llevar a cabo estas integraciones en términos de funciones elementales.

Si el objeto del cálculo integral fuera integrar funciones explícitamente, se habría llegado a un alto definitivo. Sin embargo, tal restringido objetivo no posee ninguna justificación intrínseca; es de una naturaleza artificial.

Se sabe que la integral de toda función continua *existe* como un límite y es ella misma una función continua del límite superior, bien sea que la integral pueda ser integrada en términos de funciones elementales o no. Los rasgos distintivos de las funciones elementales están basados en el hecho de que sus propiedades son fácilmente reconocidas, que su aplicación a problemas numéricos es facilitada mediante tablas convenientes, o bien que pueden ser calculadas con un grado tan grande de precisión como se desee.

Siempre que la integral de una función no pueda ser expresada por medio de funciones con las cuales ya se tenga familiaridad, no existe objeción para introducir esta integral como una nueva función "superior", lo que no es otra cosa que darle a la integral un nombre. El que la introducción de tal nueva función sea conveniente depende de las propiedades que posee, la frecuencia con la que aparece y la facilidad con la cual puede ser manipulada en la teoría y en la práctica. En este sentido el proceso de integración es un principio general para la generación de nuevas funciones.

Cierta familiaridad con este principio se ha obtenido ya al tratar con las funciones elementales. Así, se tuvo necesidad de introducir (p. 167) la integral de $1/x$ como una nueva función, la cual se denominó logaritmo y cuyas propiedades pudieron fácilmente ser deducidas. Las funciones trigonométricas pudieron haberse introducido del mismo modo, haciendo uso solamente de las funciones racionales, el proceso de integración y la operación de inversión. Con este objeto, se requiere tomar solamente una u otra de las ecuaciones

$$\operatorname{arc\,tan} x = \int_0^x \frac{dt}{1+t^2},$$

o

$$\operatorname{arc\,sen} x = \int_0^x \frac{dt}{\sqrt{1-t^2}},$$

como la *definición* de la función arc tan x, o bien arc sen x, respectivamente, y obtener entonces por inversión las funciones trigonométricas. Mediante este proceso la definición de estas funciones está divorciada de la intuición geométrica (en particular, de la noción intuitiva de "ángulo"), pero queda entonces la tarea de desarrollar estas propiedades independientemente de la geometría.[1] (Posteriormente, en la sección 3.16 se dará otra discusión puramente analítica de las funciones trigonométricas.)

*Integrales elípticas

El primer ejemplo importante que rebasa los límites del conjunto de las funciones elementales está dado por las *integrales elípticas*. Estas son

[1] No continuaremos aquí con el desarrollo de estas ideas. El paso esencial es el de probar los teoremas de adición para las funciones inversas, esto es, para el seno y para la tangente.

integrales en las que el integrando depende racionalmente de la raíz cuadrada de un polinomio de tercer o cuarto grado. Entre estas integrales, la función

$$u(s) = \int_0^s \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}$$

ha llegado a ser particularmente importante. Su función inversa $s(u)$ también desempeña un papel importante.[2] Esta función $s(u)$ ha sido examinada detenidamente y tabulada, como se hizo con las funciones elementales.[3]

Es el prototipo de las llamadas *funciones elípticas,* que tienen un lugar principalísimo en la teoría de funciones de variable compleja y aparecen en muchas aplicaciones físicas (por ejemplo, en conexión con el movimiento de un péndulo simple; ver p. 429).

El nombre "integral elíptica" surge del hecho de que tales integrales aparecen en el problema de determinar la longitud de arco de una elipse (Ver capítulo 4, pp. 395-396).

Debe hacerse notar además que integrales que a primera vista tienen una apariencia diferente resultan ser integrales elípticas después de una substitución sencilla. Como un ejemplo, la integral

$$\int \frac{dx}{\sqrt{\cos\alpha - \cos x}}$$

es transformada por medio de la substitución $u = \cos x/2$ en la integral

$$-k\sqrt{2}\int \frac{du}{\sqrt{(1-u^2)(1-k^2u^2)}}, \qquad k = \frac{1}{\cos(\alpha/2)};$$

y la integral

$$\int \frac{dx}{\sqrt{\cos 2x}}$$

se convierte por medio de la substitución $u = \operatorname{sen} x$ en

$$\int \frac{du}{\sqrt{(1-u^2)(1-2u^2)}},$$

y, finalmente, la integral

$$\int \frac{dx}{\sqrt{1-k^2\operatorname{sen}^2 x}}$$

es transformada por medio de la substitución $u = \operatorname{sen} x$ en

$$\int \frac{du}{\sqrt{(1-u^2)(1-k^2u^2)}}.$$

---

[2] Para el valor especial $k = 0$ se obtiene $u(s) = \operatorname{arc\,sen} x$ y $s(u) = \operatorname{sen} u$ respectivamente.

[3] La función $s(u)$, una de las llamadas *funciones elípticas jacobianas,* es denotada usualmente por el símbolo $\operatorname{sn} u$ para indicar que es una generalización de la función seno ordinaria.

b. Sobre la derivación y la integración

Debe agregarse otro comentario sobre la relación entre derivación e integración. La derivación puede ser considerada como un proceso más elemental que la integración, puesto que ella no conduce fuera del dominio de las funciones "conocidas". Por otra parte, debe recordarse que la derivabilidad de una función continua, arbitraria, no es de ninguna manera una conclusión predeterminada sino una hipótesis estipulada. En efecto, como se ha visto, existen funciones continuas que no son derivables en ciertos puntos aislados; en tanto que desde los tiempos de Weierstrass han sido construidos muchos ejemplos de funciones continuas que no poseen derivada en ninguna parte.[1] En contraste, aun cuando la integración en términos de funciones elementales no es posible en general, es cierto, al menos, que la integral de una función continua existe.

Tomadas en conjunto, la integración y la derivación no pueden ser comparadas entre sí sencillamente como operaciones más elementales o menos elementales una respecto a la otra; desde algunos puntos de vista, la primera puede ser considerada como más elemental; desde otros puntos de vista, la segunda.

En cuanto a lo que concierne al concepto de integral, no se utilizará en la siguiente sección la hipótesis de que el integrando es continuo en todas partes; y se verá que la integral puede ser extendida a una amplia clase de funciones que posee discontinuidades.

3.15 Extensión del concepto de integral

a. Introducción. Definición de integrales "impropias"

En el capítulo 2 (p. 150) se definió $\int_a^b f(x)\,dx$ formando las "sumas de Riemann",

$$F_n = \sum_{i=1}^{n} f(\xi_i)\,\Delta x_i,$$

basadas en la subdivisión del intervalo $[a, b]$ en n subintervalos de longitudes Δx_i y en una elección de puntos intermedios ξ_i de esos subintervalos. Si las sucesiones F_n tienden al mismo límite, F_a^b, para cualquier sucesión de subdivisiones y puntos intermedios, cuando el valor Δx_i máximo tiende a cero, se define $\int_a^b f(x)\,dx$ como el límite F_a^b. Se mostró que este límite existe

[1] Compárese con Titchmarsh, *The Theory of Functions,* Oxford, 1932, Secciones 11.21 a 11.23, pp. 350-354.

cuando $f(x)$ es continua en $[a, b]$. Sin embargo, a menudo se presenta la necesidad de definir una integral cuando $f(x)$ no está definida o no es continua *en todos* los puntos del intervalo cerrado I, o bien cuando el intervalo de integración se extiende a infinito. Sería deseable, por ejemplo, adjudicar un significado apropiado a expresiones tales como

$$\int_0^1 \frac{1}{\sqrt{x}}\, dx, \qquad \int_0^1 \operatorname{sen} \frac{1}{x}\, dx,$$

$$\int_0^\infty e^{-x}\, dx, \qquad \int_1^\infty \frac{\operatorname{sen} x}{x^2}\, dx, \qquad \text{etc.}$$

En primer lugar, se extiende el concepto de integral a funciones que son continuas en el intervalo abierto (a, b) pero que no están necesariamente definidas o son continuas en los puntos extremos a, b. Para números cualesquiera α, β, con $a < \alpha < \beta < b$, la integral ordinaria ("propia") $\int_\alpha^\beta f(x)\, dx$ queda entonces definida. Si

$$F = \lim_{\varepsilon \to 0} \int_{\alpha_\varepsilon}^{\beta_\varepsilon} f(x)\, dx$$

existe cuando $a < \alpha_\varepsilon < \beta_\varepsilon < b$ y $\lim_{\varepsilon \to 0} \alpha_\varepsilon = a$, $\lim_{\varepsilon \to 0} \beta_\varepsilon = b$, y si F es independiente de la elección particular de α_ε y β_ε, se dice que la integral impropia $\int_a^b f(x)\, dx$ converge y posee el valor F.

Integrandos seccionalmente continuos. Si, más generalmente, $f(x)$ está definida y es continua en (a, b) con la posible excepción de un número finito de puntos intermedios $c_1, c_2, \ldots, c_n$, y si f es continua en cada uno de los intervalos abiertos (a, c_1), (c_1, c_2), $\ldots$, (c_n, b) se define $\int_a^b f(x)\, dx$ como la suma de integrales impropias sobre los subintervalos, siempre que cada una de éstas converja.

La integral impropia $\int_a^b f(x)\, dx$ siempre converge cuando f es continua y *acotada* en un intervalo abierto (a, b). Por ejemplo, la integral

$$\int_0^1 \operatorname{sen} \frac{1}{x}\, dx = \lim_{\varepsilon \to 0} \int_\varepsilon^1 \operatorname{sen} \frac{1}{x}\, dx$$

converge. Para probar esta propiedad general puede suponerse, por brevedad, que f es continua en b, pero no necesariamente en a. Entonces, por definición,

$$\int_a^b f(x)\, dx = \lim_{\alpha \to a} F(\alpha),$$

donde $F(\alpha)$, para $a < \alpha < b$, está definida como $\int_\alpha^b f(x)\,dx$. Si M es una cota superior para f y α_n es una sucesión que tiende a a, se tiene, por el teorema del valor medio del cálculo integral, $|F(\alpha_n) - F(\alpha_m)| \leq M\,|\alpha_n - \alpha_m|$; y, consecuentemente, por el criterio de convergencia de Cauchy, $\lim\limits_{n\to\infty} F(\alpha_n)$ existe. Puesto que éste es el caso para cualquier sucesión α_n que converge a a, se sigue que $\lim\limits_{\alpha\to a} F(\alpha)$ existe.

De hecho, cuando f es continua y acotada en (a,b) se pueden asignar a f cualesquiera valores en los puntos extremos a, b y obtener también $\int_a^b f(x)\,dx$ directamente como una integral "propia" definida como el límite de sumas de Riemann. Se ve fácilmente que para funciones continuas y acotadas f ambas definiciones se aplican y conducen al mismo valor, independientemente de la elección de $f(a)$ y $f(b)$. Lo mismo es cierto, más generalmente, para funciones acotadas que están definidas y son continuas en (a, b) con la posible excepción de un número finito de puntos. En particular, $\int_a^b f(x)\,dx$ siempre existe cuando f es continua excepto en un

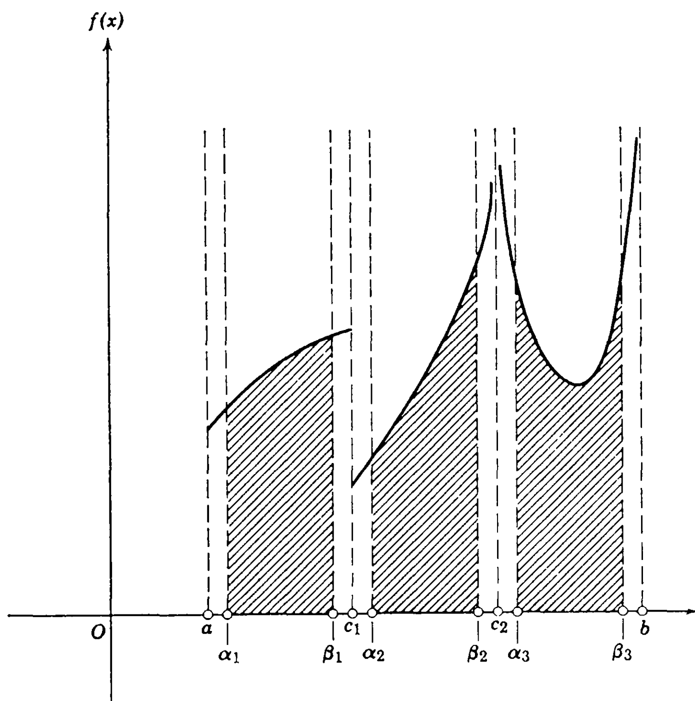


Figura 3.30 La integral de una función con discontinuidades.

número finito de discontinuidades de salto. Por todo lo anterior, la convergencia de una integral impropia de una función sobre un intervalo finito requiere atención solamente cuando f se torna infinita.

Nótese que la interpretación geométrica de la integral como el área bajo la curva es la misma que para una f continua (Fig. 3.30).

b. Funciones con discontinuidades infinitas

Se comenzará con la integral

$$J = \int_0^1 \frac{dx}{x^\alpha},$$

donde α es un número positivo. El integrando $1/x^\alpha$ se torna infinito para $x \to 0$. Por lo tanto, se debe definir J tomando la integral J_ε desde el límite positivo ε hasta el límite 1, y, finalmente, hacer que ε tienda a cero. De acuerdo con las reglas elementales de integración se obtiene, suponiendo $\alpha \neq 1$,

$$J_\varepsilon = \int_\varepsilon^1 \frac{dx}{x^\alpha} = \frac{1}{1-\alpha}(1 - \varepsilon^{1-\alpha}).$$

Se observan de inmediato las siguientes posibilidades: 1) α es mayor que 1; entonces, para $\varepsilon \to 0$ el miembro derecho tiende a infinito. 2) α es menor que 1; entonces el miembro derecho tiende al límite $1/(1-\alpha)$. En el segundo caso, por lo tanto, solamente se tiene que tomar este valor límite como la integral $J = \int_0^1 \frac{dx}{x^\alpha}$. En el primer caso se dice que la integral desde 0 hasta 1 *no existe*, o bien, que *diverge*. 3) En el tercer caso, en el que $\alpha = 1$, la integral es igual a $-\log \varepsilon$ y, por lo tanto, para $\varepsilon \to 0$ no tiende a un límite sino que tiende a infinito; esto es, la integral $\int_0^1 \frac{dx}{x} = J$ no existe, o sea, es divergente.

Otro ejemplo de un integrando con una discontinuidad infinita está dado por $f(x) = 1/\sqrt{1-x^2}$. Se encuentra

$$\int_0^{1-\varepsilon} \frac{dx}{\sqrt{1-x^2}} = \text{arc sen}\,(1-\varepsilon).$$

Para $\varepsilon \to 0$ el miembro derecho converge al límite $\pi/2$; éste es, entonces, el valor de la integral:

$$\frac{\pi}{2} = \int_0^1 \frac{dx}{\sqrt{1-x^2}},$$

aun cuando el integrando se torna infinito en el punto $x = 1$.

c. Interpretación de la integral como un área

Las integrales impropias pueden ser interpretadas como áreas de regiones que se extienden al infinito definidas a partir de regiones acotadas, por medio del paso al límite. Por ejemplo, los resultados anteriores para la función $1/x^\alpha$ establecen que el área acotada por el eje x, la recta $x = 1$, la recta $x = \varepsilon$ y la curva $y = 1/x^\alpha$

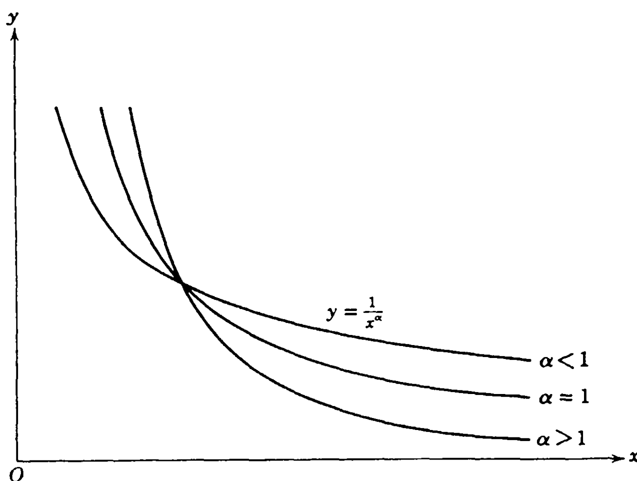


Figura 3.31 Para ilustrar la convergencia o la divergencia de integrales impropias.

tiende a un límite finito cuando $\varepsilon \to 0$, siempre que $\alpha < 1$, y que tiende a infinito si $\alpha \geq 1$. Este hecho puede ser expresado como sigue: El área entre el eje x, el eje y, la curva $y = 1/x^\alpha$ y la recta $x = 1$ es finita o infinita según que $\alpha < 1$ o $\alpha \geq 1$.

La intuición no puede dar, por supuesto, una información segura sobre la finitud o la infinitud del área de una región que se extiende al infinito. La figura 3.31 ilustra el hecho de que para $\alpha < 1$ el área bajo la curva permanece finita, en tanto que para $\alpha \geq 1$ es infinita; un hecho que, ciertamente, no es sugerido por la intuición geométrica.

d. Criterios para determinar la convergencia

Para probar la convergencia de la integral de una función $f(x)$ con una discontinuidad infinita en el punto $x = b$, puede hacerse uso a menudo del siguiente criterio.

Sea la función $f(x)$ continua en el intervalo $a < x < b$, y supóngase que $\lim_{x \to b} f(x) = \infty$. Entonces, la integral $\int_a^b f(x)\,dx$ *converge* si existe un número positivo μ menor que uno y un número fijo M, independientes de x, tal que en todas partes del intervalo $a < x < b$ se cumple la desigualdad $|f(x)| \leq M/(b - x)^\mu$; en en otras palabras, *si en el punto $x = b$ la función $f(x)$ se torna infinita de orden menor que el primero:* $f(x) = O[1/(b - x)^\mu]$ para algún $\mu < 1$. Por otra parte, la integral *diverge* si existen un número $\nu \geq 1$ y un número fijo N tales que en todas partes del intervalo $a \leq x < b$ se cumple la desigualdad $f(x) \geq N/(b - x)^\nu$; en otras palabras, *si en el punto $x = b$ la función positiva $f(x)$ se torna infinita de primer orden al menos.*

La demostración surge casi de inmediato por comparación con el caso muy sencillo que se acaba de discutir. Para probar la primera parte del teorema se observa que para $0 < \varepsilon < b - a$ se tiene

$$0 \leq \frac{M}{(b - x)^\mu} + f(x) \leq \frac{2M}{(b - x)^\mu},$$

y, por lo tanto, también

$$0 \leq \int_a^{b-\varepsilon} \left[\frac{M}{(b-x)^\mu} + f(x) \right] dx \leq \int_a^{b-\varepsilon} \frac{2M}{(b-x)^\mu} dx.$$

Si $\varepsilon \to 0$, el integrando en la derecha, el cual se obtiene a partir de la integral $\int dx/x^\mu$ mediante una sencilla substitución de x por $b - x$, posee un límite, y, por ende, permanece acotado. Además, los valores de la integral de en medio crecen monótonamente conforme $\varepsilon \to 0$; puesto que también están acotados, deben poseer un límite, y así la integral

$$\int_a^b \left(\frac{M}{(b-x)^\mu} dx + f(x) \right) dx$$

$$= \lim_{\varepsilon \to 0} \left(\int_a^{b-\varepsilon} \frac{2M}{(b-x)^\mu} dx + \int_a^{b-\varepsilon} f(x)\, dx \right)$$

converge. La convergencia de la integral de $M/(b-x)^\mu$ implica entonces también la de $\int_a^b f(x) dx$.

La demostración de la segunda parte del teorema se deja como un ejercicio para el lector.

De la misma manera se ve de inmediato que teoremas exactamente análogos se cumplen cuando el límite o extremo *inferior* de integración es un punto de discontinuidad infinita. Si un punto de discontinuidad infinita está situado en el interior del intervalo de integración, se separa simplemente el intervalo en dos subintervalos por medio de este punto y se aplican entonces estas consideraciones a cada uno de esos subintervalos.

Como un ejemplo se considera la integral elíptica

$$\int_0^1 \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}} \qquad (k^2 < 1).$$

Por la identidad $1 - x^2 = (1-x)(1+x)$ se ve de inmediato que conforme $x \to 1$ el integrando se torna infinito, pero sólo de orden $\frac{1}{2}$; de lo que se sigue que la integral impropia converge. (Para $k = 1$ la integral diverge.)

e. Intervalo de integración infinito

Otra extensión importante del concepto de integral se refiere a la de un intervalo de integración infinito. Para una formulación precisa se introduce la siguiente notación: Si la integral

$$\int_a^A f(x)\, dx,$$

con a fijo, tiende a un límite definido cuando $A \to \infty$, la integral de $f(x)$ sobre el intervalo infinito $x \geq a$ se define como

$$\lim_{A \to \infty} \int_a^A f(x)\, dx = \int_a^\infty f(x)\, dx.$$

Otra vez, tal integral se denomina *convergente*.

Ejemplos. Ejemplos sencillos de las diversas posibilidades están dados aquí también por las funciones $f(x) = 1/x^\alpha$:

$$\int_1^A \frac{dx}{x^\alpha} = \frac{1}{1-\alpha}(A^{1-\alpha} - 1).$$

Aquí se ve, una vez más, que si se excluye el caso $\alpha = 1$, la integral extendida al infinito existe para el caso $\alpha > 1$, y, en efecto,

$$\int_1^\infty \frac{dx}{x^\alpha} = \frac{1}{\alpha - 1}.$$

Cuando $\alpha < 1$ la integral ya no existe. Para el caso $\alpha = 1$, claramente la integral de nuevo deja de existir puesto que $\log x$ tiende a infinito cuando x tiende a infinito. Por ello se ve que, en lo que respecta a la integración sobre un intervalo infinito, las funciones $1/x^\alpha$ no se comportan del mismo modo que para la integración hasta el origen. Esta afirmación se hace también plausible mediante una ojeada a la Fig. 3.31. Pues, obviamente, mientras más grande sea α las curvas se acercarán más rápidamente al eje x para $x \to \infty$; y así es plausible que el área considerada tienda a un límite definido para valores suficientemente grandes de α.

El siguiente criterio para la existencia de una integral con un límite infinito es a menudo útil. (Se supone nuevamente que para valores suficientemente grandes de x, digamos, para $x \geq a$, el integrando es continuo.)

Criterio de convergencia

La integral $\int_a^\infty f(x)\,dx$ *converge si la función* $f(x)$ *se anula en el infinito con un orden superior al primero;* esto es, si existe un número $\nu > 1$ tal que para todos los valores de x suficientemente grandes se cumple la relación $|f(x)| \leq M/x^\nu$, donde M es un número fijo independiente de x. En símbolos: $f(x) = 0\left(\frac{1}{x^\nu}\right)$ Por otra parte, *la integral diverge si la función permanece positiva y se anula en el infinito con un orden no superior al primero,* esto es, si existe un número fijo $N > 0$ tal que $xf(x) \geq N$.

La demostración de estos criterios es exactamente paralela al argumento anterior y puede dejarse al lector.

Un ejemplo muy sencillo es la integral $\int_a^\infty \frac{1}{x^2}\,dx$ $(a > 0)$. El integrando se anula en el infinito como infinitésimo de segundo orden. Se ve de inmediato que la integral converge, pues $\int_a^A \frac{dx}{x^2}\,dx = \frac{1}{a} - \frac{1}{A}$, y, por ende,

$$\int_a^\infty \frac{1}{x^2}\,dx = \frac{1}{a}.$$

Otro ejemplo igualmente simple es

$$\int_0^\infty \frac{1}{1+x^2}\,dx = \lim_{A\to\infty}(\arctan A - \arctan 0) = \frac{\pi}{2}.$$

Entonces, evidentemente también se tiene

$$\int_{-\infty}^{+\infty} \frac{1}{1+x^2}\,dx = \pi,$$

puesto que el integrando es una función par. Es curioso que el área entre la curva $y = 1/(1 + x^2)$ y el eje x (ver Fig. 3.8, p. 238), que se extiende al infinito, resulte ser la misma que la del círculo de radio uno.

f. La función gamma

Otro ejemplo de particular importancia en el análisis es el de la llamada función gamma:

$$\Gamma(n) = \int_0^\infty e^{-x}x^{n-1}\,dx \qquad (n > 0).$$

Partiendo el intervalo de integración en dos partes, una desde $x = 0$ hasta $x = 1$ y otra desde $x = 1$ hasta $x = \infty$, se ve que la integral sobre la primera parte converge, puesto que $0 < e^{-x}x^{n-1} < 1/x^\mu$ con $\mu = 1 - n < 1$. Para la integral sobre la segunda, la parte infinita, el criterio de convergencia también se satisface; por ejemplo, para $\nu = 2$ se tiene $\lim_{x\to\infty} x^2e^{-x}x^{n-1} = 0$, puesto que la función exponencial e^{-x} tiende a cero con un orden superior al de cualquier potencia $1/x^m$ $(m > 0)$ (ver p. 274). Esta función gamma, que se considera como una función del número n (no necesariamente un entero), satisface una relación notable que se obtiene mediante integración por partes como sigue. Primero se tiene (con $f(x) = x^{n-1}$, $g'(x) = e^{-x}$)

$$\int e^{-x}x^{n-1}\,dx = -e^{-x}x^{n-1} + (n-1)\int e^{-x}x^{n-2}\,dx.$$

Si se toma esta relación integral entre 0 y A y entonces se deja crecer A más allá de toda cota, se obtiene inmediatamente

$$\Gamma(n) = (n-1)\int_0^\infty e^{-x}x^{n-2}\,dx = (n-1)\Gamma(n-1) \qquad \text{para} \quad n > 1;$$

y de esta fórmula de recurrencia, siempre que μ sea un entero y $0 < \mu < n$, se sigue que

$$\Gamma(n) = (n-1)(n-2)\cdots(n-\mu)\int_0^\infty e^{-x}x^{n-\mu-1}\,dx.$$

En particular, si n es un entero positivo se tiene, con $\mu = n - 1$, la relación

$$\Gamma(n) = (n-1)(n-2)\cdots 3\cdot 2\cdot 1\int_0^\infty e^{-x}\,dx;$$

y, puesto que

$$\int_0^\infty e^{-x}\,dx = 1,$$

se tiene finalmente

$$\Gamma(n) = (n-1)(n-2)\cdots 2\cdot 1 = (n-1)!,$$

una expresión muy útil de un factorial mediante una integral.

Otros ejemplos. Las integrales

$$\int_0^\infty e^{-x^2}\,dx, \qquad \int_0^\infty x^n e^{-x^2}\,dx$$

también convergen, como puede deducirse fácilmente a partir del criterio antes dado. La primera es idéntica con $\frac{1}{2}\Gamma(\frac{1}{2})$, la segunda con $\frac{1}{2}\Gamma[(n+1)/2]$ para $n > -1$, como se ve mediante la substitución $x^2 = u$, $dx = (1/2\sqrt{u})\,du$.

g. La integral de Dirichlet

En muchas aplicaciones se encuentran integrales cuya convergencia no surge directamente a partir de nuestro criterio. Un ejemplo importante es proporcionado por la integral

$$I = \int_0^\infty \frac{\operatorname{sen} x}{x}\,dx$$

investigada por Dirichlet. Si el límite superior no es infinito sino finito, la integral es convergente puesto que la función $(\operatorname{sen} x)/x$ es continua para todo x finito (para $x = 0$, está dada por $\lim \dfrac{\operatorname{sen} x}{x} = 1$ con $x \to 0$). La convergencia de la integral I se debe al cambio periódico del signo del integrando, lo que hace que las contribuciones a la integral debidas a los intervalos vecinos de longitud π casi se cancelen unas con otras (Fig. 3.32). Así, la suma del número infinito de áreas entre el eje x

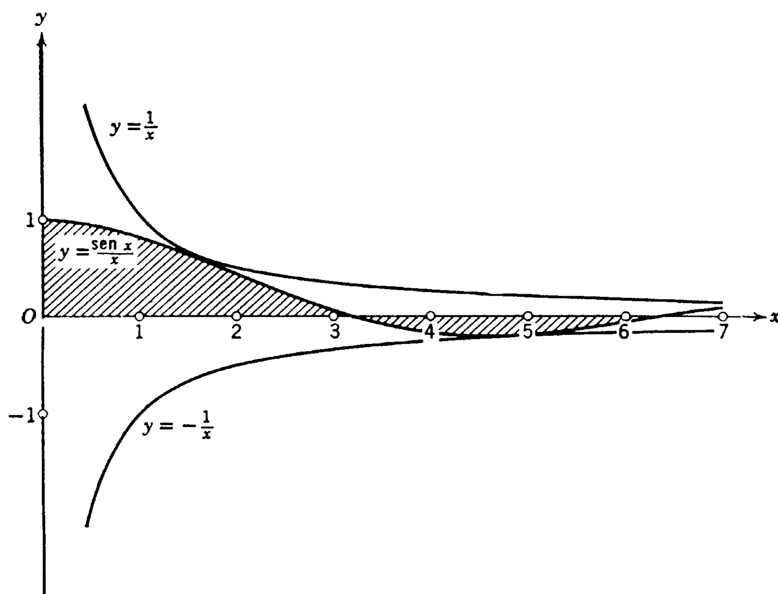


Figura 3.32 Gráfica de $y = \dfrac{\operatorname{sen} x}{x}$.

y la curva $y = \frac{\operatorname{sen} x}{x}$ converge, si se cuentan las áreas por encima del eje x como positivas y por debajo del eje x como negativas. (Por otra parte, la suma de los valores numéricos de todas las áreas, esto es, la integral

$$\int_0^\infty \frac{|\operatorname{sen} x|}{x}\, dx$$

como puede fácilmente mostrarse, diverge.)

El carácter alternante de la función sen x es causa del hecho de que su integral indefinida

$$\int \operatorname{sen} x\, dx = 1 - \cos x$$

está acotada para todo x. Se hace uso de esto al calcular el valor de la expresión

$$I_{AB} = \int_A^B \frac{\operatorname{sen} x}{x}\, dx = \int_A^B \frac{1}{x} \frac{d(1 - \cos x)}{dx}\, dx$$

La integración por partes muestra que

$$I_{AB} = \frac{1 - \cos B}{B} - \frac{1 - \cos A}{A} + \int_A^B \frac{1 - \cos x}{x^2}\, dx.$$

Por lo tanto,

$$\int_0^\infty \frac{\operatorname{sen} x}{x}\, dx = \lim_{\substack{A \to 0 \\ B \to \infty}} I_{AB} = \int_0^\infty \frac{1 - \cos x}{x^2}\, dx,$$

donde la integral en el miembro derecho es claramente convergente. En otras palabras, la integral I existe. En la sección 8.4*c* se establecerá, además, el hecho notable de que I posee el valor $\pi/2$.

h. Substitución. Integrales de Fresnel

Evidentemente, todas las reglas para la substitución de nuevas variables, etcétera, permanecen válidas para integrales impropias convergentes. A menudo tales transformaciones pueden conducir a expresiones diferentes más manejables para la integral.

Como un ejemplo, para calcular

$$\int_0^\infty x e^{-x^2}\, dx$$

se introduce la nueva variable $u = x^2$ y se obtiene

$$\int_0^\infty x e^{-x^2}\, dx = \frac{1}{2} \int_0^\infty e^{-u}\, du = \lim_{A \to \infty} \frac{1}{2} (1 - e^{-A}) = \frac{1}{2}.$$

Otro ejemplo en la investigación de integrales impropias está dado por las *integrales de Fresnel*, que aparecen en la teoría de difracción de la luz:

$$F_1 = \int_0^\infty \operatorname{sen}(x^2)\,dx, \qquad F_2 = \int_0^\infty \cos(x^2)\,dx.$$

La substitución $x^2 = u$ conduce a

$$F_1 = \frac{1}{2}\int_0^\infty \frac{\operatorname{sen} u}{\sqrt{u}}\,du, \qquad F_2 = \frac{1}{2}\int_0^\infty \frac{\cos u}{\sqrt{u}}\,du.$$

Integrando por partes se encuentra

$$\int_A^B \frac{\operatorname{sen} u}{\sqrt{u}}\,du = \frac{1-\cos B}{\sqrt{B}} - \frac{1-\cos A}{\sqrt{A}} + \frac{1}{2}\int_A^B \frac{1-\cos u}{u^{3/2}}\,du.$$

Si A y B tienden a cero e infinito respectivamente, se ve, mediante el mismo argumento que para la integral de Dirichlet, que la integral F_1 converge. La convergencia de la integral F_2 se prueba exactamente de la misma manera.

Estas integrales de Fresnel muestran que una integral impropia puede existir aún si el integrando no tiende a cero cuando $x \to \infty$. En efecto, una integral impropia puede existir aún cuando el integrando no esté acotado, como se muestra mediante el ejemplo

$$\int_0^\infty 2u\cos(u^4)\,du.$$

Cuando $u^4 = n\pi$, esto es, cuando $u = \sqrt[4]{n\pi}$, $n = 0, 1, 2, \ldots$, el integrando se convierte en $2\sqrt[4]{n\pi}\cos n\pi = \pm 2\sqrt[4]{n\pi}$, de modo que no es acotado. Mediante la substitución $u^2 = x$, sin embargo, la integral se reduce a

$$\int_0^\infty \cos(x^2)\,dx,$$

la cual, justamente, se ha visto que es convergente.

Por medio de una substitución una integral impropia puede ser a menudo transformada en una integral propia. Por ejemplo, la transformación $x = \operatorname{sen} u$ conduce a

$$\int_0^1 \frac{dx}{\sqrt{1-x^2}} = \int_0^{\pi/2} du = \frac{\pi}{2}.$$

Por otra parte, integrales de funciones continuas pueden ser transformadas en integrales impropias. Esto sucede si la transformación $u = \phi(x)$ es tal que en un extremo del intervalo de integración la derivada $\phi'(x)$ se anula, de modo que dx/du es infinita.

3.16 Las ecuaciones diferenciales de las funciones trigonométricas

a. Comentarios introductorios sobre las ecuaciones diferenciales

La integración es simplemente un primer paso en un campo mucho más extenso: En vez de invertir la derivación mediante la integración, esto es, resolver la ecuación $y' = f(x)$ con una $f(x)$ dada, para tener $y = F(x)$, puede desearse encontrar funciones $y = F(x)$ que satisfagan relaciones más generales entre y y las derivadas de y. Tales "ecuaciones diferenciales" aparecen en todas partes en las aplicaciones, así como en contextos estrictamente teóricos. Sobre estas ecuaciones se hacen estudios penetrantes que van mucho más allá del alcance de este libro. Se volverá sobre algunos aspectos elementales de la teoría de las ecuaciones diferenciales posteriormente, en éste y en el siguiente volumen. Por ahora la discusión se restringirá a ejemplos bastante sencillos pero significativos. Se discutirán las ecuaciones diferenciales de las funciones $\operatorname{sen} x$ y $\cos x$, las que han sido mencionadas ya en la p. 192.

Aunque en la trigonometría elemental estas funciones y sus propiedades fueron tomadas desde el punto de vista geométrico, se descarta ahora la confianza en la intuición geométrica y se expresan las funciones trigonométricas en una forma sencilla sobre una base precisa, analítica, de acuerdo con la tendencia general del desarrollo mencionado anteriormente.

b. Sen x y cos x definidos mediante una ecuación diferencial y condiciones iniciales

Se considera la ecuación diferencial

$$u'' + u = 0$$

con el deseo de caracterizar soluciones $u(x)$ que serán identificadas con las funciones seno y coseno. Cualquier función $u = F(x)$ que satisface la ecuación, esto es, para la cual $F''(x) + F(x) = 0$, se denomina una *solución*.[1]

Salta a la vista que, junto con una solución $u = F(x)$, la función $u = F(x + h)$, para una constante arbitraria h, es también una solución, como se verifica de inmediato derivando $F(x + h)$ dos veces con respecto a x. Análogamente, se ve de inmediato que, conjuntamente con $F(x)$, la derivada $F'(x) = u$ es también una solución; como también lo es, claro está, $cF(x)$, con el factor constante c. Además, junto con $F_1(x)$ y $F_2(x)$, cualquier combinación lineal $c_1F_1(x) + c_2F_2(x) = F(x)$, con constantes c_1 y c_2, es una solución.

[1] Por supuesto, siempre se entiende que las funciones consideradas son derivables un número suficiente de veces.

Para destacar de la multitud de soluciones de una ecuación diferencial una específica se imponen "condiciones iniciales" estipulando que para $x = 0$ los valores de $u = F(0)$ y $u' = F'(0)$ estén prefijados como a y b respectivamente. Se establece primero el siguiente *teorema:*

La solución está determinada de manera única por estos valores iniciales.

Para la demostración, se comienza con un comentario general válido para cualquier solución u. Multiplicando la ecuación diferencial por $2u'$ se encuentra, debido a que $2u''u' = (u'^2)'$ y $2u'u = (u^2)'$, la ecuación

$$0 = 2u''u' + 2u'u = [(u')^2 + u^2]',$$

la cual puede ser integrada de inmediato e implica

$$u'^2 + u^2 = c,$$

donde c es una constante, esto es, no depende de x. Por lo tanto, c debe poseer el mismo valor que el miembro izquierdo para $x = 0$. Así se tiene que para cualquier solución u es

$$u'^2(0) + u^2(0) = c.$$

Ahora supóngase que se tienen dos soluciones u_1 y u_2 con las mismas condiciones iniciales: Entonces la diferencia $z = u_1 - u_2$ es una solución con $z'(0) = z(0) = 0$. Por lo tanto, se tiene $c = 0$ y para todo x: $x'^2 + z^2 = 0$. Esto significa que $z = 0$ y $z' = 0$, lo que demuestra el aserto.

Se definen ahora las funciones $\operatorname{sen} x$ y $\cos x$ como aquellas soluciones de la ecuación diferencial $u''(x) + u(x) = 0$ para las cuales las condiciones iniciales son, respectivamente, para $u = \operatorname{sen} x$,

$$u(0) = a = 0, \qquad u'(0) = b = 1,$$

y para $u = \cos x$,

$$u(0) = a = 1, \qquad u'(0) = b = 0.$$

Se toma por cierto aquí que tales soluciones existen y son arbitrariamente derivables cuantas veces se quiera; mientras que la demostración de esto se dará posteriormente en un contexto más general (ver sección 9.2).[1]

La única solución u de $u'' + u = 0$ para la cual es $u = a$, $u' = b$ para $x = 0$ es entonces la función $u = a \cos x + b \operatorname{sen} x$. Esto prueba que toda solución de la ecuación diferencial es una combinación lineal de $\cos x$ y $\operatorname{sen} x$.

Se obtienen ahora las propiedades básicas de las funciones trigonométricas a partir de la ecuación diferencial $u'' + u = 0$ aplicada, por ejemplo,

---

[1] Obsérvese que estos hechos pueden inferirse inmediatamente a partir de la ecuación $u'^2 + u^2 = 1$, la cual es válida para $\operatorname{sen} x$, así como también para $\cos x$, y, a partir de su forma equivalente $dx/du = 1/\sqrt{1 - u^2}$, las funciones inversas de $\operatorname{sen} x$ y $\cos x$ son obtenidas inmediatamente por integración.

a la función $u = \text{sen}\, x$. Obviamente, si u es una solución, $v = u'$ también lo es: $v'' + v = 0$. Debido a que $u'' + u = v' + u = 0$, se tiene $v'(0) = -u(0) = 0$, mientras que $v(0) = u'(0) = 1$. Por lo tanto,

$$v(x) = \cos x = \frac{d}{dx} \text{sen}\, x.$$

Análogamente se deduce $(d/dx) \cos x = -\text{sen}\, x$.

El teorema central de la trigonometría es el teorema de adición:

$$\cos (x + y) = \cos x \cos y - \text{sen}\, x \,\text{sen}\, y.$$

Se sigue de inmediato, a partir del presente enfoque, que: en primer lugar, la función $\cos (x + y)$, como una función de x, manteniendo por el momento a y constante, es una solución $u(x)$ de la ecuación diferencial $u'' + u = 0$, que satisface para $x = 0$ las condiciones iniciales $u(0) = \cos y = a$ y $u'(0) = -\text{sen}\, y = b$; en segundo lugar, la solución —que, de acuerdo con el teorema anterior, es la única— para la cual $u(0) = a$ y $u'(0) = b$, es precisamente $a \cos x + b \,\text{sen}\, x$, como se verifica de inmediato. Por lo tanto, se tiene de inmediato para nuestra solución $\cos (x + y)$ la expresión

$$\cos (x + y) = \cos x \cos y - \text{sen}\, x \,\text{sen}\, y,$$

como se quería probar.

Los comentarios en esta sección podrían ser suficientes para indicar cómo las funciones trigonométricas pueden ser introducidas de una manera enteramente analítica sin referencia alguna a la geometría.

Sin entrar en más detalles, se menciona lo siguiente.

El número $\frac{1}{2}\pi$ podría ser definido ahora como el menor valor positivo de x para el cual $\cos x = 0$.

La periodicidad de las funciones trigonométricas surge asimismo fácilmente a partir del enfoque analítico.

Se volverá a la construcción analítica de las funciones trigonométricas mediante las series infinitas de potencias (ver sección 5.56).

PROBLEMAS

SECCION 3.1, página 223

1. Sea $P(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$.

(*a*) Calcular el polinomio $F(x)$ a partir de la ecuación

$$F(x) - F'(x) = P(x).$$

*(*b*) Calcular $F(x)$ a partir de la ecuación

$$c_0F(x) + c_1F'(x) + c_2F''(x) = P(x).$$

2. Encontrar el límite, para $n \to \infty$, del valor absoluto de la n-ésima derivada de $1/x$ en el punto $x = 2$.

3. Probar que si $f^{(n)}(x) = 0$ para todo x, entonces f es un polinomio de grado $n - 1$ a lo más, y recíprocamente.

4. Determinar la forma de una función racional r para la cual

$$\lim_{x\to\infty} \frac{xr'(x)}{r(x)} = 0.$$

5. Probar por inducción que la n-ésima derivada de un producto puede encontrarse de acuerdo con la siguiente regla (regla de Leibnitz):

$$\frac{d^n}{dx^n}(fg) = f\frac{d^n g}{dx^n} + \binom{n}{1}\frac{df}{dx}\frac{d^{n-1}g}{dx^{n-1}} + \binom{n}{2}\frac{d^2f}{dx^2}\frac{d^{n-2}g}{dx^{n-2}} + \cdots$$
$$+ \binom{n}{n-1}\frac{d^{n-1}f}{dx^{n-1}}\frac{dg}{dx} + \frac{d^nf}{dx^n}g.$$

Aquí $\binom{n}{1} = n$, $\binom{n}{2} = \frac{n(n-1)}{2!}$, etc., denotan los coeficientes binomiales.

6. Probar que $\sum_{i=1}^{n-1} ix^{i-1} = \frac{(n-1)x^n - nx^{n-1} + 1}{(x-1)^2}$.

SECCION 3.2, página 228

1. Sea $y = e^x(a \operatorname{sen} x + b \cos x)$. Demostrar que y'' puede expresarse como una combinación lineal de y e y', esto es,

$$y'' = py' + qy,$$

donde p y q son constantes. Expresar todas las derivadas superiores como combinaciones lineales de y' e y.

*2. Encontrar la n-ésima derivada de arc sen x en $x = 0$, y luego la de (arc sen $x)^2$ en $x = 0$.

SECCION 3.3, página 239

1. Encontrar la segunda derivada de $f[g\{h(x)\}]$.

2. Derivar la siguiente función: $\log_{v(x)} u(x)$, [esto es, el logaritmo de $u(x)$ en la base $v(x)$; y sea $v(x) > 0$].

3. ¿Qué condiciones deben satisfacer los coeficientes α, β, a, b, c para que

$$\frac{\alpha x + \beta}{\sqrt{(ax^2 + 2bx + c)}}$$

posea en todas partes una derivada finita que en ningún caso se anule?

4. Mostrar que $d^n(e^{x^2/2})/dx^n = u_n(x)e^{x^2/2}$, donde $u_n(x)$ es un polinomio de grado n. Establecer la relación de recurrencia

$$u_{n+1} = xu_n + u_n'.$$

*5. Mediante la aplicación de la regla de Leibnitz a

$$\frac{d}{dx}(e^{x^2/2}) = xe^{x^2/2},$$

obténgase la relación de recurrencia

$$u_{n+1} = xu_n + nu_{n-1}.$$

***6.** Mediante la combinación de las relaciones de recurrencia de los problemas 4 y 5, obtener la ecuación diferencial

$$u_n'' + xu_n' - nu_n = 0$$

satisfecha por $u_n(x)$.

7. Encontrar la solución polinomial

$$u_n(x) = x^n + a_1x^{n-1} + \cdots + a_n.$$

de la ecuación diferencial $u_n'' + xu_n' - nu_n = 0$.

***8.** Si $P_n(x) = \dfrac{1}{2^n n!}\dfrac{d^n}{dx^n}(x^2 - 1)^n$, probar las relaciones

(a) $P'_{n+1} = \dfrac{x^2 - 1}{2(n + 1)}P_n'' + \dfrac{(n + 2)x}{n + 1}P_n' + \dfrac{n + 2}{2}P_n.$

(b) $P'_{n+1} = xP_n' + (n + 1)P_n.$

(c) $\dfrac{d}{dx}[(x^2 - 1)P_n'] - n(n + 1)P_n = 0.$

9. Encontrar la solución polinomial

$$P_n = \frac{(2n)!}{2^n(n!)^2}x^n + a_1x^{n-1} + \cdots + a_n$$

de la ecuación diferencial

$$\frac{d}{dx}[(x^2 - 1)P_n'] - n(n + 1)P_n = 0.$$

10. Determinar el polinomio $P_n(x) = \dfrac{1}{2^n n!}\dfrac{d^n}{dx^n}(x^2 - 1)^n$ utilizando el teorema del binomio.

***11.** Sea $\lambda_{n,p}(x) = \dbinom{p}{n}x^n(1 - x)^{p-n}$, $n = 0, 1, 2, \ldots, p$. Mostrar que

$$1 = \sum_{n=0}^{p} \lambda_{n,p}(x).$$

$$x^k = \sum_{n=k}^{p} \frac{\dbinom{n}{k}}{\dbinom{p}{k}} \lambda_{n,p}(x).$$

. . . .

$$x^p = \lambda_{p,p}(x).$$

. . . .

SECCION 3.4, página 244

1. La función $f(x)$ satisface la ecuación

$$f(x + y) = f(x)f(y).$$

$a)$ Si $f(x)$ es derivable, entonces $f(x) \equiv 0$ o bien $f(x) = e^{ax}$.
$*b)$ Si $f(x)$ es continua, entonces $f(x) \equiv 0$ ó bien $f(x) = e^{ax}$.

2. Si una función derivable $f(x)$ satisface la ecuación

$$f(xy) = f(x) + f(y),$$

entonces $f(x) = \alpha \log x$.

3. Probar que si $f(x)$ es continua y

$$f(x) = \int_0^x f(t)\, dt,$$

entonces $f(x)$ es idénticamente cero.

SECCION 3.5, página 250

1. Demostrar la fórmula

$$\operatorname{senh} a + \operatorname{senh} b = 2 \operatorname{senh} \left(\frac{a+b}{2} \right) \cosh \left(\frac{a-b}{2} \right).$$

Obténganse fórmulas análogas para $\operatorname{sen} a - \operatorname{senh} b$, $\cosh a + \cosh b$, $\cosh a - \cosh b$.

2. Expresar $\tanh (a + b)$ en términos de $\tanh a$ y de $\tanh b$.
Expresar $\coth (a + b)$ en términos de $\coth a$ y $\coth b$.
Expresar sen $\frac{1}{2}a$ y cosh $\frac{1}{2}a$ en términos de $\cosh a$.

3. Derivar:
(a) $\cosh x + \operatorname{senh} x$; (b) $e^{\tanh x + \coth x}$,
(c) $\log \operatorname{senh} (x + \cosh^2 x)$; (d) $\operatorname{arc\,cosh} x + \operatorname{arc\,senh} x$,
(e) $\operatorname{arc\,senh} (\alpha \cos x)$; (f) $(\operatorname{arc\,tanh} (2x/(1 + x^2)))$.

4. Calcular el área acotada por la catenaria $y = \cosh x$, las ordenadas $x = a$ y $x = b$ y el eje x.

SECCION 3.6, página 257

1. Determinar los máximos, mínimos y puntos de inflexión de $x^3 + 3px + q$. Discutir la naturaleza de las raíces de $x^3 + 3px + q = 0$.

2. Dada la parábola $y^2 = 2px$, $p > 0$, y un punto $P(x = \xi, y = \eta)$ dentro de ella $(\eta^2 < 2p\xi)$, encontrar la trayectoria mínima (consistente en dos segmentos de recta) que va de P a un punto Q sobre la parábola y de ahí al foco $F(x = \frac{1}{2}p, y = 0)$ de la parábola. Demostrar que el ángulo FQP es bisectado por la normal a la parábola y que QP es paralelo al eje de la parábola (principio del espejo parabólico).

3. De todos los triángulos con base y ángulo vertical dados, el triángulo isósceles posee el área máxima.

4. De todos los triángulos con base y área dadas, el triángulo isósceles posee el máximo ángulo vertical.

*5. De todos los triángulos con área dada, el triángulo equilátero posee el mínimo perímetro.

*6. De todos los triángulos con perímetro dado, el triángulo equilátero posee el área máxima.

*7. De todos los triángulos inscritos en una circunferencia, el triángulo equilátero posee el área máxima.

8. Probar que si $p > 1$ y $x > 0$, es $x^p - 1 \geq p(x - 1)$.

9. Probar la desigualdad $1 > (\operatorname{sen} x)/x \geq 2/\pi$, $0 \leq x \leq \pi/2$.

10. Probar que: (a) $\tan x \geq x$, $0 \leq \pi/2$;
(b) $\cos x \geq 1 - x^2/2$.

*11. Dados $a_1 > 0$, $a_2 > 0, \ldots, a_n > 0$, determínese el mínimo de

$$\frac{\dfrac{a_1 + \cdots + a_{n-1} + x}{n}}{\sqrt[n]{a_1 a_2 \cdots a_{n-1} x}}$$

para $x > 0$. Utilizar el resultado para probar por inducción matemática que (cfr. problema 13, p. 131):

$$\sqrt[n]{a_1 a_2 \cdots a_n} \leq \frac{a_1 + \cdots + a_n}{n}.$$

12. (*a*) Dados n números fijos $a_1, \ldots, a_n$, determinar x de modo que $\sum_{i=1}^{n} (a_i - x)^2$ sea un mínimo.

*(*b*) Minimizar $\sum_{i=1}^{n} |a_i - x|$.

*(*c*) Minimizar $\sum_{i=1}^{n} \lambda_i |a_i - x|$. donde $\lambda_i > 0$.

13. Bosquejar la gráfica de la función

$$y = (x^2)^x, \; y(0) = 1.$$

Demuéstrese que la función es continua en $x = 0$. ¿Posee la función máximos, mínimos o puntos de inflexión?

*14. Encontrar el mínimo valor α tal que

$$\left(1 + \frac{1}{x}\right)^{x+\alpha} > e$$

para todo x positivo. (*Sugerencia:* Es sabido que $[1 + (1/x)]^{x+1}$ decrece y $[1 + (1/x)]^x$ crece, ambas monótonamente, al límite e en el infinito.)

*15. (*a*) Encontrar el punto para el cual la suma de sus distancias a los tres lados de un triángulo es un mínimo.

(*b*) Encontrar el punto para el cual la suma de sus distancias a los vértices es un mínimo.

16. Probar las siguientes desigualdades:

(*a*) $e^x > 1/(1 + x), x > 0$.

(*b*) $e^x > 1 + \log(1 + x), x > 0$.

(*c*) $e^x > 1 + (1 + x) \log(1 + x), x > 0$.

17. Supóngase que $f''(x) < 0$ en (a, b). Probar que:

(*a*) Todo arco de la gráfica dentro del intervalo está situado por encima de la cuerda que une sus puntos extremos.

(*b*) La gráfica está situada por debajo de la tangente, para cualquier punto dentro de (a, b).

*18. Sea f una función que tiene segunda derivada en (a, b).

(*a*) Demostrar que cualquiera de las condiciones *a* o *b* del problema 22 es suficiente para garantizar que $f''(x) \leq 0$.

(*b*) Demostrar que la condición

$$f\left(\frac{x + y}{2}\right) \geq \frac{f(x) + f(y)}{2}$$

para todo x e y en (a, b) es suficiente para garantizar que $f''(x) \leq 0$.

*19. Sean a, b, dos números positivos, y p y q números cualesquiera diferentes de cero y tales que $p < q$. Probar que

$$\frac{[\theta a^p + (1-\theta)b^p]^{1/p}}{[\theta a^q + (1-\theta)b^q]^{1/q}} \leq 1$$

para todos los valores de θ en el intervalo $0 < \theta < 1$.

(Esta es la desigualdad de Jensen, que establece que la potencia media p-ésima $[\theta a^p + (1-\theta)b^p]^{1/p}$ de dos cantidades a, b es una función creciente de p.)

20. Mostrar que el signo de igualdad en la desigualdad anterior se cumple si y sólo si $a = b$.

21. Probar que $\lim_{p\to 0} [\theta a^p + (1-\theta)b^p]^{1/p} = a^{\theta}b^{1-\theta}$.

22. Definiendo la potencia media 0-ésima de a, b, como $a^{\theta}b^{1-\theta}$, demostrar que la desigualdad de Jensen se aplica a este caso y se convierte en (con $a \neq b$):

$$a^{\theta}b^{1-\theta} \gtrless [\theta a^q + (1-\theta)b^q]^{1/q} \text{ de acuerdo con que } q \gtrless 0.$$

Para $q = 1$, es

$$a^{\theta}b^{1-\theta} \leq \theta a + (1-\theta)b.$$

23. Probar la desigualdad

$$a^{\theta}b^{1-\theta} \leq \theta a + (1-\theta)b,$$

$a, b > 0$, $0 < \theta < 1$, sin referencia a la desigualdad de Jensen, y mostrar que la igualdad se cumple sólo si $a = b$. (Esta igualdad establece que la media geométrica de θ, $1-\theta$, es menor que la correspondiente media aritmética.)

***24.** Sea f continua y positiva en $[a, b]$ y denótese por M su valor máximo. Probar que

$$M = \lim_{n\to\infty} \sqrt[n]{\int_a^b [f(x)]^n\, dx}.$$

SECCION 3.7, página 269

1. Sea $f(x)$ una función continua que se anula, junto con su primera derivada, para $x = 0$. Mostrar que $f(x)$ se anula con un orden superior al de x cuando $x \to 0$.

2. Mostrar que

$$f(x) = \frac{a_0x^n + a_1x^{n-1} + \cdots + a_n}{b_0x^m + b_1x^{m-1} + \cdots + b_m},$$

donde a_0, $b_0 \neq 0$, es del mismo orden de magnitud que x^{n-m} cuando $x \to \infty$.

***3.** Probar que e^x no es una función racional.

***4.** Probar que e^x no puede satisfacer una ecuación algebraica con polinomios en x como coeficientes.

5. Si el orden de magnitud de la función positiva $f(x)$ cuando $x \to \infty$ es mayor, el mismo o bien menor que el de x^m, probar que $\int_a^x f(\xi)\,d\xi$ posee el correspondiente orden de magnitud relativo a x^{m+1}.

6. Comparar el orden de magnitud, para $x \to \infty$, de $\int_a^x f(\xi)\,d\xi$ relativo a $f(x)$ para las siguientes funciones $f(x)$:

(*a*) $\dfrac{e^{\sqrt{x}}}{\sqrt{x}}$.   (*c*) xe^{x^2}.

(*b*) e^x.   (*d*) $\log x$.

SECCION 3.8, página 284

1. Encontrar el límite, para $n \to \infty$, de $a_n = \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n}$.

*2. Encontrar el límite de

$$b_n = \frac{1}{\sqrt{n^2 - 0}} + \frac{1}{\sqrt{n^2 - 1}} + \frac{1}{\sqrt{n^2 - 4}} + \cdots + \frac{1}{\sqrt{n^2 - (n-1)^2}}.$$

*3. Si α es cualquier número real mayor que -1, evaluar

$$\lim_{n\to\infty} \frac{1^\alpha + 2^\alpha + 3^\alpha + \cdots + n^2}{n^{\alpha+1}}.$$

SECCION 3.11, página 295

1. Demostrar que para todos los valores positivos impares de n, la integral $\int e^{-x^2}x^n\,dx$ puede ser evaluada en términos de funciones elementales.

2. Demostrar que si n es par, la integral $\int e^{-x^2}x^n\,dx$ puede ser evaluada en términos de funciones elementales y de la integral $\int e^{-x^2}\,dx$ (para la cual se han construido tablas).

3. Probar que

$$\int_0^x \left[\int_0^u f(t)\,dt\right]du = \int_0^x f(u)(x-u)\,du.$$

*4. El problema 3 da una fórmula para la segunda integral iterada. Probar que la n-ésima integral iterada de $f(x)$ está dada por

$$\frac{1}{(n-1)!}\int_0^x f(u)(x-u)^{n-1}\,du.$$

5. Probar para el coeficiente binomial $\binom{y}{u}$ que

$$\binom{n}{k} = \left[(n+1)\int_0^1 x^k(1-x)^{n-k}\,dx\right]^{-1}$$

6. Obtener una fórmula recursiva para

$$\int x^p(ax^n + b)^q\,dx$$

y utilizar esta relación para integrar

$$\int x^3(x^7+1)^4\,dx.$$

*7. (*a*) Sea $P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n$: Demostrar que

$$\int_{-1}^1 P_n(x)P_m(x)\,dx = 0, \quad \text{si } m \neq \nu.$$

(*b*) Probar que $\int_{-1}^1 P_n{}^2(x)\,dx = \frac{2}{2n+1}$.

(*c*) Probar que $\int_1^{-1} x^m P_n(x)\,dx = 0$, si $m < n$.

(*d*) Evaluar $\int_{-1}^{1} x^n P_n(x)\,dx$.

SECCION 3.12, página 302

*1. Integrar

$$\int \frac{dx}{x^6 + 1}.$$

2. Utilizar el desarrollo en fracciones parciales para probar las fórmulas de Newton:

$$\frac{\alpha_1{}^k}{g'(\alpha_1)} + \frac{\alpha_2{}^k}{g'(\alpha_2)} + \cdots + \frac{\alpha_n{}^k}{g'(\alpha_n)} = \begin{cases} 0 & \text{para } k = 0, 1, 2, \ldots, n-2 \\ 1 & \text{para } k = n-1, \end{cases}$$

donde $g(x)$ es un polinomio de la forma $x^n + \alpha_1 x^{n-1} + \cdots$ con raíces distintas $\alpha_1, \ldots, \alpha_n$.

SECCION 3.14, página 317

*1. Probar que la substitución $x = (\alpha t + \beta)/(\gamma t + \delta)$, con $\alpha\delta - \gamma\beta \neq 0$, transforma la integral

$$\int \frac{dx}{\sqrt{ax^4 + bx^3 + cx^2 + dx + e}}$$

en una integral de un tipo similar, y que si la expresión bicuadrática

$$ax^4 + bx^3 + cx^2 + dx + e$$

no posee factores repetidos, tampoco los posee la nueva expresión bicuadrática en t que toma su lugar. Probar que lo mismo se cumple para

$$\int R(x, \sqrt{ax^4 + bx^3 + cx^2 + dx + e})\,dx,$$

donde R es una función racional.

2. La función

$$\phi(x) = \int_0^x \frac{du}{\sqrt{1 - k^2 \operatorname{sen}^2 u}}$$

es conocida como la *integral elíptica de primer orden*.

(*a*) Mostrar que ϕ es continua y creciente, y que, por lo tanto, posee una inversa continua.

(*b*) Denótese por $am(x)$ la inversa de $\phi(x)$. Probar que $sn(x) = \operatorname{sen}[am(x)]$, donde $sn(x)$ está definida en la p. 319, nota al pie número 3.

SECION 3.15, página 320

*1. Probar que $\int_0^\infty \operatorname{sen}^2\left[\pi\left(x + \frac{1}{x}\right)\right] dx$ *no existe* (diverge).

*2. Probar que $\lim_{k\to\infty} \int_0^\infty \frac{dx}{1 + kx^{10}} = 0$.

3. ¿Para qué valores de s es (*a*) $\int_0^\infty \frac{x^{s-1}}{1+x}\,dx$, (*b*) $\int_0^\infty \frac{\operatorname{sen} x}{x^s}\,dx$ convergente?

*4. ¿Converge la integral $\int_0^\infty \frac{\operatorname{sen} t}{1+t}\,dt$?

*5. (*a*) Si a es un número positivo fijo, probar que

$$\lim_{h\to 0+} \int_{-a}^{a} \frac{h}{h^2+x^2}\,dx = \pi.$$

(*b*) Si $f(x)$ es continua en el intervalo $-1 \le x \le 1$, probar que

$$\lim_{h\to 0} \int_{-1}^{1} \frac{h}{h^2+x^2} f(x)\,dx = \pi f(0).$$

*6. Probar que $\lim_{x\to\infty} e^{-x^2}\int_0^x e^{t^2}\,dt = 0.$

7. Suponiendo que $|\alpha| \ne |\beta|$, probar que

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \operatorname{sen} \alpha x \operatorname{sen} \beta x\,dx = 0.$$

*8. Si la integral $\int_a^\infty \frac{f(x)}{x}\,dx$ converge para cualquier valor positivo a, y si $f(x)$ tiende a un límite L cuando $x \to 0$, mostrar que $\int_0^\infty \frac{f(\alpha x) - f(\beta x)}{x}\,dx$ converge para α y β positivos y posee el valor $L \log \frac{\beta}{\alpha}$.

9. Refiriéndose al problema 8, probar que

(*b*) $$\int_0^\infty \frac{e^{-\alpha x} - e^{-\beta x}}{x}\,dx = \log \frac{\beta}{\alpha}.$$

(*a*) $$\int_0^\infty \frac{\cos \alpha x - \cos \beta x}{x}\,dx = \log \frac{\beta}{\alpha}.$$

*10. Si la integral $\int_a^b \frac{f(x)}{x}\,dx$ converge para cualesquiera valores positivos de a y b, y si $f(x)$ tiende a un límite M cuando $x \to \infty$ y a un límite L cuando $x \to 0$, mostrar que

$$\int_0^\infty \frac{f(\alpha x) - f(\beta x)}{x}\,dx = (L - M)\log \frac{\beta}{\alpha}.$$

11. Obténganse las expresiones siguientes para la función gamma:

$$\Gamma(n) = 2\int_0^\infty x^{2n-1} e^{-x^2}\,dx,$$

$$\Gamma(n) = \int_0^1 \left(\log \frac{1}{x}\right)^{n-1} dx.$$

SECCION 3.16, página 331

1. Obtener la fórmula de adición para sen $(x + y)$.

2. Sin utilizar las fórmulas de adición, probar que $\cos x$ es una función par y en x una función impar.

3. (*a*)* Probar que para algún h positivo es $\cos x < 1$ para $0 < x < h$.
(*b*) Si $\cos z > 0$ para $0 \leq z \leq 2^n x$, probar que

$$\cos (2^{n+1}x) < 2^n (\cos x - 1) + 1.$$

(*c*) Combinando los resultados (*a*) y (*b*), probar que $\cos x$ posee un cero.

4. Sea a el menor cero positivo de $\cos x$. Probar que

$$\operatorname{sen} (x + 4a) = \operatorname{sen} x,$$

$$\cos (x + 4a) = \cos x.$$

5. Completar en los pasos siguientes la demostración indirecta de que $\cos x$ posee un cero:
(*a*) Si $\cos x$ no posee ceros, entonces $\operatorname{sen} x$ es monótona creciente para $x \geq 0$.
(*b*) Las funciones $\operatorname{sen} x$ y $\cos x$ están acotadas por arriba y por abajo.
(*c*) El límite de $\operatorname{sen} x$ cuando x tiende a infinito existe y es positivo.
(*d*) La ecuación

$$\cos x = 1 - \int_0^x \operatorname{sen} t \, dt$$

está en contradicción con el inciso (*b*).

PROBLEMAS DIVERSOS

1. Probar que

$$\frac{d^n}{dx^n} f(\log x) = x^{-n} \frac{d}{dt} \left(\frac{d}{dt} - 1 \right) \left(\frac{d}{dt} - 2 \right) \cdots \left(\frac{d}{dt} - n + 1 \right) f(t)$$

donde $t = \log x$. Aquí se emplea

$$\left(\frac{d}{dt} - k \right) \phi = \frac{d\phi}{dt} - k\phi,$$

donde ϕ es cualquier función de t y k es una constante.

2. Una curva suave cerrada C se dice que es *convexa* si está situada por completo de un lado de cada tangente. Demostrar que en el triángulo de área mínima circunscrito a C cada lado es tangente a C en el punto medio del lado.

4

Aplicaciones en física y geometría

4.1 Teoría de curvas planas

a. Representación paramétrica

Definición

La representación de una curva por medio de una ecuación $y = f(x)$ impone una seria restricción geométrica: Una curva representada de esa manera no debe ser intersectada en más de un punto por cualquier paralela al eje y. Usualmente esta restricción puede superarse descomponiendo la curva en porciones, cada una representable en la forma $y = f(x)$. Así, una circunferencia de radio a alrededor del origen está dada por las dos funciones $y = \sqrt{a^2 - x^2}$ e $y = -\sqrt{a^2 - x^2}$ definidas para $-a \leq x \leq a$. Sin embargo, para una curva tan simple como una recta paralela al eje y este artificio no sirve.

Mayor flexibilidad se obtiene usando una representación *implícita* mediante una ecuación $\phi(x, y) = 0$ que involucra una función ϕ de dos variables independientes. Por ejemplo, la circunferencia de radio a alrededor del origen está completamente descripta por $\phi(x, y) = x^2 + y^2 - a^2 = 0$. Cualquier línea recta en el plano tiene una ecuación implícita de la forma $ax + by + c = 0$, donde a, b y c son constantes y a y b no se anulan simultáneamente. Para $b = 0$ se obtiene una paralela al eje y.

La descripción implícita de una curva tiene la desventaja de que para encontrar puntos (x, y) de la curva explícitamente, es decir, el valor de y para un valor de x dado, se debe *resolver* la ecuación $\phi(x, y) = 0$. Este problema se discutirá en detalle en el Volumen II.

La descripción más directa y más flexible de una curva es una *representación paramétrica*. En vez de considerar una de las coordenadas rectangulares en particular, y o x, como una función de la otra, se consideran ambas coordenadas x e y como funciones de una *tercera* variable indepen-

diente t, el llamado *parámetro* [1]; y el punto con coordenadas x e y describe entonces la curva conforme t recorre un cierto intervalo. Tal representación paramétrica ya ha sido encontrada antes; por ejemplo, la circunferencia $x^2 + y^2 = a^2$ tiene la representación paramétrica $x = a \cos t$, $y = a \operatorname{sen} t$. Aquí t denota el ángulo en el centro de la circunferencia.

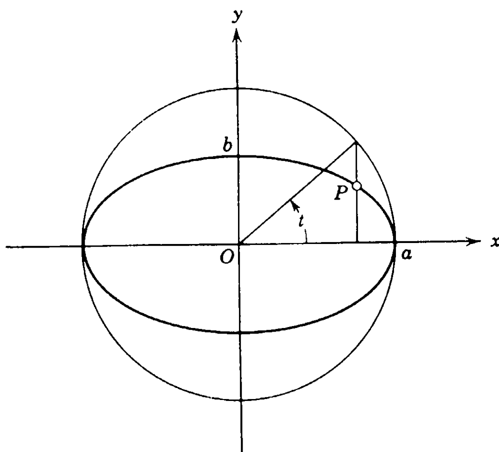


Figura 4.1

Para el elipse $x^2/a^2 + y^2/b^2 = 1$ se tiene una representación paramétrica similar: $x = a \cos t$, $y = b \operatorname{sen} t$, donde t es el llamado ángulo excéntrico, esto es, el ángulo en el centro de la elipse asociado al punto de la circunferencia circunscrita que está verticalmente por encima o por debajo del punto $P = (a \cos t, b \operatorname{sen} t)$ de la elipse. Se supone aquí que $b < a$ (véase Fig. 4.1). En ambos casos el punto con coordenadas x, y describe por completo la circunferencia o la elipse conforme el parámetro t recorre el intervalo $0 \leq t < 2\pi$.

En general, una curva C es representada paramétricamente mediante dos funciones de un parámetro t,

$$x = \phi(t) = x(t), \qquad y = \psi(t) = y(t);$$

la notación abreviada $x(t)$ e $y(t)$ se usará siempre que no dé lugar a confusión.*

En lo que sigue se supondrá que ϕ y ψ tienen derivadas continuas, a menos que se diga lo contrario

[1] Esta palabra denota una variable auxiliar que no se desea señalar en especial.

* La notación $x = \phi(t)$, etc., pone de relieve la conexión funcional entre la variable dependiente y la independiente; la notación $x(t)$, etc., significa justamente que t va a ser considerada como la variable independiente que determina el valor funcional de x según una forma prescrita.

Transformación del intervalo paramétrico en la curva. Sentido de dirección

Dada una curva, las dos funciones $\phi(t)$ y $\psi(t)$ deben determinarse de manera que el conjunto de pares de valores funcionales, $x(t)$ e $y(t)$, correspondientes a un cierto intervalo de valores de t definan todos los puntos de la curva y ningún otro punto más. Se tiene entonces una correspondencia entre los puntos de la curva y los valores de t en un intervalo del eje t. La representación paramétrica define una *transformación de un intervalo del eje t en la curva,*[2] siendo el punto t original del eje t transformado en el punto $x = \phi(t)$, $y = \psi(t)$ de C.

Puesto que $x(t)$ e $y(t)$ se han considerado continuas, puntos vecinos sobre el eje t corresponden a puntos vecinos sobre la curva. Además, como los puntos del eje t están ordenados, de manera obvia se puede asignar un orden o "sentido" a los puntos de C diciendo simplemente que el punto en el cual se transforma el número t_1 *precede* al punto en el cual se transforma t_2 si $t_1 < t_2$ (ver p. 352). La representación paramétrica da así un significado preciso a la vaga noción intuitiva de curva como un conjunto de puntos dispuestos en el mismo orden que sobre una línea recta.

b. Cambios de parámetros

Los valores del parámetro t sirven para distinguir los diferentes puntos de la curva C. Dichos valores desempeñan el papel de "nombres" para los puntos individuales de la curva.

La misma curva C admite muchas representaciones paramétricas diferentes. Cualquier cantidad que varíe continuamente a lo largo de la curva y que tenga diferentes valores en puntos diferentes de ella pueed servir como parámetro.

Si, por ejemplo, la curva está dada originalmente por una ecuación $y = f(x)$, se puede escoger como parámetro t la variable x y describir la curva mediante las funciones $x = t$, $y = f(t)$. Análogamente, para una curva descripta dando x como una función de y, por ejemplo $x = g(y)$, se puede usar y como parámetro t y escribir $x = g(t)$, $y = t$.

Para una curva dada por una ecuación $r = h(\theta)$ en *coordenadas polares* r, θ (véase capítulo 1, p. 124), se puede escoger θ como parámetro t y obtener la representación paramétrica

$$x = r\cos\theta = h(t)\cos t = \phi(t),$$

$$y = r\operatorname{sen}\theta = h(t)\operatorname{sen} t = \psi(t).$$

[2] Más exactamente, se dice que la transformación es *sobre* la curva C (en inglés, *"onto"*) porque los puntos imagen son *todos* los puntos de la curva *(N. del R.)*.

Partiendo de una representación paramétrica dada, $x = \phi(t)$, $y = \psi(t)$, para una curva C, siempre se pueden obtener muchas otras representaciones paramétricas. Para ello se toma una función arbitraria $\tau = \chi(t)$ que sea monótona y continua en el intervalo t correspondiente a los puntos de C. La función χ tiene entonces una inversa monótona y continua $t = \sigma(\tau)$ en un correspondiente intervalo de τ. Las coordenadas de los puntos (x, y) de C pueden entonces representarse en la forma

$$x = \phi[\sigma(\tau)] = \alpha(\tau), \quad y = \psi[\sigma(\tau)] = \beta(\tau).$$

Las funciones $\alpha(\tau)$ y $\beta(\tau)$ son también continuas; además, puntos diferentes de C corresponden a valores diferentes de t y, por lo tanto, debido al carácter monótono de la función σ, a diferentes valores de τ. El efecto total del cambio de parámetro de t a τ es el de "rebautizar" los puntos de C.

Así, la recta $y = x$ tiene la representación paramétrica $x = t$, $y = t$, donde $-\infty < t < \infty$. La substitución $\tau = t^3$ da lugar a la representación paramétrica $x = \tau^{1/3}$, $y = \tau^{1/3}$ para la misma recta.

Análogamente, la elipse $x^2/a^2 + y^2/b^2 = 1$ admite la representación paramétrica $x = a \cos t$, $y = b \operatorname{sen} t$, donde $0 \leq t < 2\pi$. Definiendo $t = c\zeta + d$, para números reales c y d $(c \neq 0)$, se obtiene otra representación, $x(\zeta) = a \cos(c\zeta + d)$, $y(\zeta) = b \operatorname{sen}(c\zeta + d)$, para la misma elipse, con ζ variando en los intervalos $-d/c \leq \zeta < (2\pi - d)/c$ para $c > 0$ y $(2\pi - d)/c < \zeta \leq -d/c$ para $c < 0$. La substitución $\tau = \tan(t/2)$ conduce a la *representación paramétrica "racional"* (ver pp. 311-312),

$$x = \frac{a(1 - \tau^2)}{1 + \tau^2}, \quad y = \frac{2b\tau}{1 + \tau^2},$$

para la elipse. Conforme τ recorre todos los valores reales, se obtienen todos los puntos de la elipse con la excepción del punto $S = (-a, 0)$.

Las singularidades en una representación ordinaria pueden desaparecer si se usa un parámetro adecuado. Por ejemplo, la curva $y = \sqrt[3]{x^2}$ puede ser representada por las funciones suaves $x = t^3$, $y = t^2$. El punto con coordenadas x, y describe entonces la curva completa (parábola semicúbica) conforme t varía de $-\infty$ a $+\infty$.

Esta flexibilidad en la elección del parámetro permite a menudo simplificar el estudio de las propiedades geométricas que, por supuesto, no dependen de representaciones específicas.

En particular, puede resultar conveniente algunas veces el uso de una representación $y = f(x)$ para C o parte de C. Tal representación es siempre posible para un pedazo de la curva $(t_0 \leq t \leq t_1)$ en el cual una de las funciones ϕ, ψ (por ejemplo, $x = \phi(t)$), es monótona. En efecto, para este pedazo se tiene una función inversa única $t = \gamma(x)$, y así es $y = \psi[\gamma(x)]$.[1]

[1] Esto es, por supuesto, sólo una afirmación relativa a una propiedad "local" de una curva, o sea, una afirmación que se hace solamente para una parte pequeña

c. Movimiento a lo largo de una curva. El tiempo como un parámetro. El ejemplo de la cicloide

Movimiento a lo largo de una curva

Muy a menudo el parámetro t tiene el natural significado físico de tiempo. Cualquier movimiento de un punto en el plano puede describirse representando sus coordenadas x e y como funciones del tiempo de manera que al tiempo t el punto (x, y) esté en $(x(t), y(t))$. Por consiguiente, estas dos funciones determinan en forma paramétrica el movimiento a lo largo de un camino o trayectoria C; y ellas constituyen una *transformación* de la escala del tiempo sobre la trayectoria.[2]

Las cicloides y las trocoides

Un ejemplo lo proporcionan las *cicloides,* o sea, las trayectorias de puntos sobre un círculo que rueda uniformemente sin deslizar a lo largo de una línea recta u otro círculo. En el caso más simple un círculo de radio a rueda a lo largo del eje x; y la trayectoria de un punto P sobre su circunferencia es una cicloide "común". Escójanse el origen de coordenadas del sistema y el tiempo inicial de manera que para el tiempo $t = 0$ el punto P esté en el origen y que en el tiempo t el círculo haya girado un cierto ángulo t a partir de su posición original. Esto significa que el círculo gira en el sentido de las manecillas de un reloj "con velocidad angular" igual a uno. Se supone que el círculo rueda uniformemente sin deslizar a lo largo del eje x, de manera que para el tiempo t la distancia del punto de contacto al origen es exactamente igual a la longitud de arco medida desde el punto de contacto hasta P. Así, en el tiempo t el centro M del círculo giratorio debe estar en el punto (at, a); y el centro se mueve con velocidad constante a hacia la derecha. Para las coordenadas de P en el tiempo t se encuentra entonces (ver Fig. 4.2) la representación paramétrica

$$(1) \qquad x = a(t - \operatorname{sen} t), \quad y = a(1 - \cos t).$$

Eliminando el parámetro t puede obtenerse la ecuación de la curva en forma no paramétrica a costa, sin embargo, de la claridad de la expresión. Se tiene

$$\cos t = \frac{a - y}{a}, \quad t = \operatorname{arc} \cos \frac{a - y}{a}, \quad \operatorname{sen} t = \pm \sqrt{1 - \frac{(a - y)^2}{a^2}},$$

de la curva, escogida adecuadamente. Generalmente (por ejemplo, en el caso de la circunferencia), la variable x no puede usarse como parámetro para la curva completa sino sólo para una porción de ella.

[2] A un cambio del parámetro t correspondería entonces un cambio en la escala del tiempo según la cual la curva queda descripta por el punto móvil.

y, por lo tanto,

$$x = a \arccos \frac{a-y}{a} \mp \sqrt{y(2a-y)}, \tag{1a}$$

obteniéndose así x como una función de y.

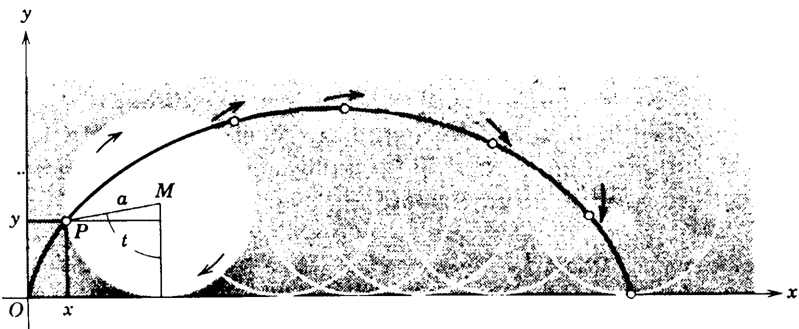


Figura 4.2 Cicloide.

Epicicloide

Nuestro siguiente ejemplo es el de una *epicicloide,* definida como la trayectoria de un punto P fijo sobre la circunferencia de un círculo de radio c, conforme éste rueda a una velocidad uniforme a lo largo de la circunferencia y fuera de un segundo círculo de radio a. Supóngase el círculo fijo centrado en el origen del plano x, y. Supóngase además que el círculo que se mueve rueda a lo largo del círculo fijo de manera que su centro ha girado alrededor del origen a un ángulo t en el tiempo t (Fig. 4.3). Entonces se encuentra que en el tiempo t la posición del punto $P = (x(t), y(t))$, el cual en el tiempo $t = 0$ es el punto de contacto $(a, 0)$, está dada por las ecuaciones paramétricas

$$\begin{aligned} x(t) &= (a+c)\cos t - c\cos\left(\frac{a+c}{c}t\right), \\ y(t) &= (a+c)\operatorname{sen} t - c\operatorname{sen}\left(\frac{a+c}{c}t\right). \end{aligned} \tag{2}$$

Cuando $a = c$ la curva formada se llama *cardioide* (Fig. 4.4) y está dada por las ecuaciones paramétricas

$$\begin{aligned} x(t) &= 2a\cos t - a\cos(2t), \\ y(t) &= 2a\operatorname{sen} t - a\operatorname{sen}(2t). \end{aligned} \tag{3}$$

Una tercera variedad de cicloides se obtiene como el lugar geométrico de un punto sujeto a la circunferencia de un círculo que rueda sobre la circunferencia de otro círculo fijo, pero por su interior. Para encontrar las ecuaciones paramétricas de esta "*hipocicloide*", sea a el radio del círculo fijo y c el del círculo que gira. Supóngase que el punto P de la circunferencia del círculo que se mueve está localizado en $(a, 0)$ en el tiempo $t = 0$. Supóngase, además, que el círculo que gira se mueve a lo largo del círculo fijo de manera que en el tiempo t su centro ha girado alrededor del origen en un ángulo t (Fig. 4.5). Se encuentra entonces que las ecuaciones paramétricas para la hipocicloide son

$$x(t) = (a - c)\cos t + c\cos\left(\frac{a-c}{c}t\right),$$

(4)

$$y(t) = (a - c)\operatorname{sen} t - c\operatorname{sen}\left(\frac{a-c}{c}t\right).$$

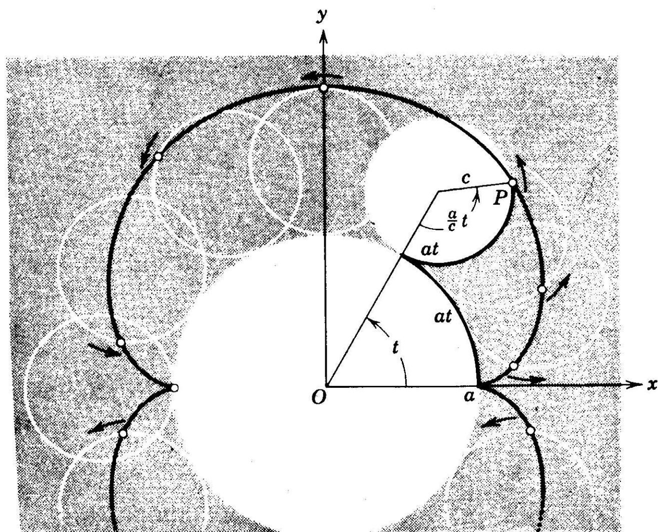


Figura 4.3 Epicicloide.

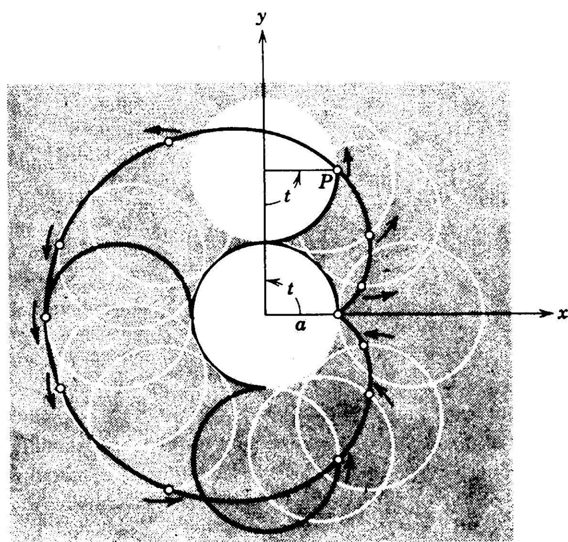


Figura 4.4 Cardioide.

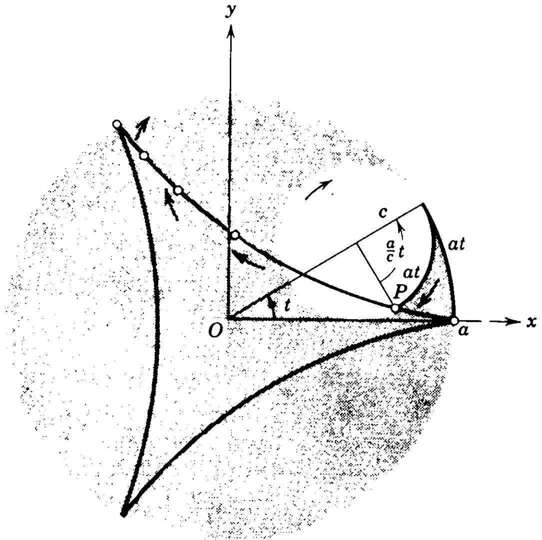


Figura 4.5 Hipocicloide.

En el caso especial en el que el círculo fijo tiene un radio igual a dos veces el radio del círculo que se mueve, $c = \frac{1}{2}a$, se obtiene

$$x(t) = a\cos t,$$
$$y(t) = 0,$$

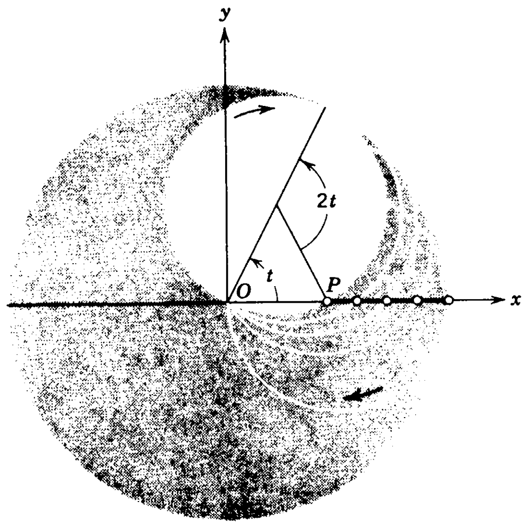


Figura 4.6 Un punto P sobre la periferia de un círculo que rueda interiormente sobre una circunferencia de radio doble describe un segmento de recta.

y la hipocicloide degenera en el diámetro del círculo fijo, recorrido hacia atrás y hacia adelante. La característica interesante de este ejemplo es que proporciona una solución mecánica al problema de dibujar una línea recta usando únicamente movimientos circulares (Fig. 4.6).

Si el radio del círculo fijo es tres veces el radio del círculo móvil, entonces $c = a/3$ y

$$x(t) = \tfrac{2}{3}a\cos t + \tfrac{1}{3}a\cos(2t),$$

$$y(t) = \tfrac{2}{3}a\,\mathrm{sen}\,t - \tfrac{1}{3}a\,\mathrm{sen}\,(2t).$$

Mediante un cálculo elemental se encuentra

$$x^2 + y^2 = \tfrac{5}{9}a^2 + \tfrac{4}{9}a^2\cos(3t),$$

de manera que la hipocicloide encuentra al círculo fijo en exactamente tres puntos y la curva aparece como se muestra en la Fig. 4.5.

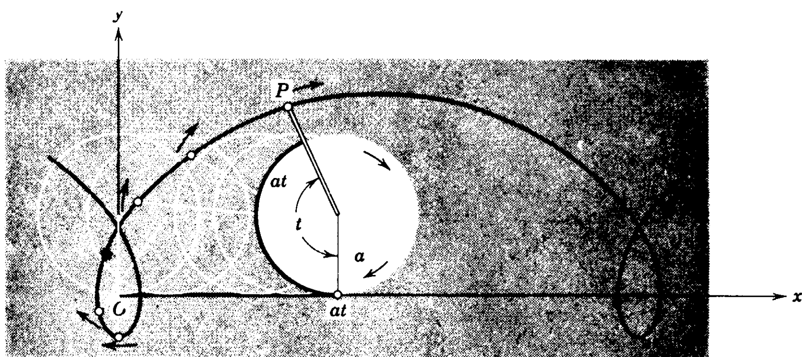


Figura 4.7 Trocoide.

Trocoides

Curvas más generales llamadas *trocoides* (epitrocoides, hipotrocoides) se obtienen al considerar el movimiento de un punto P sujeto a un círculo (pero no necesariamente sobre su periferia) cuando dicho círculo rueda a lo largo de una línea recta o a lo largo de otro círculo por su parte exterior o interior (ver Fig. 4.7). El mismo tipo de curva aparece como la trayectoria de un punto que se mueve uniformemente sobre un círculo en tanto que el centro del círculo mismo se mueve uniformemente a lo largo de una línea recta o de un círculo. Estas curvas juegan un papel muy importante en la descripción ptolomeica del movimiento aparente de los planetas.

Algunas de las propiedades notables de las cicloides serán discutidas más adelante en este capítulo (p. 446).

d. Clasificación de curvas. Orientación

Definiciones

Entre las características más obvias de una curva están el número de partes separadas o *ramas* y el número de espiras que ellas tengan. Una hipérbola es un ejemplo de una curva que consiste de dos ramas ajenas;

otro de tales ejemplos es la curva $y^2 = (4 - x^2)(x^2 - 1)$ que consiste de dos óvalos separados. Nos interesarán particularmente las curvas consistentes de una sola parte, las curvas *conexas*. Una curva conexa puede *intersectarse* a sí misma, como la trocoide (Fig. 4.7), o la "lemniscata" de la Fig. 1.S.3, p. 126.

Una curva conexa sin autointersecciones se denomina *simple*. Entre las curvas simples todavía se distinguirán las curvas *cerradas,* como los círculos o las elipses, de las curvas que no son cerradas, como parábolas o segmentos de recta. No se intentará aquí dar una clasificación ni rigurosa ni completa de las curvas, sino únicamente señalar ciertas características *"topológicas"* de una curva, importantes para la representación paramétrica.

Arcos simples

Una representación paramétrica de una curva C por medio de dos funciones continuas, $x = \phi(t)$, $y = \psi(t)$, define una transformación del eje t, o de una porción de él, en C. C se denominará un *arco simple* si puede representarse de tal manera que el parámetro t recorra un intervalo cerrado $[a, b]$ sobre el eje t, formando así el dominio de las funciones $\phi(t)$, $\psi(t)$, y si, además, diferentes t en el intervalo corresponden a diferentes puntos P sobre C. Un ejemplo es el arco parabólico $x = t$, $y = t^2$ para $0 \leq t \leq 1$.

El mismo arco C (esto es, los mismos puntos del plano) puede ser representado paramétricamente de muchas maneras. Cualquier función monótona continua $\tau = \chi(t)$, para $a \leq t \leq b$, define un parámetro τ de manera que x e y sean funciones continuas de τ en un intervalo cerrado $[\alpha, \beta]$ conveniente, correspondiendo valores diferentes de τ a diferentes P. De hecho, como es fácil concluir, las substituciones monótonas continuas $\tau = \chi(t)$ proporcionan las representaciones paramétricas continuas *más generales* de un arco simple que asignan diferentes puntos del arco a diferentes valores del parámetro. (Ver notas en la p. 78 relativas a transformaciones continuas biunívocas.)

Una representación paramétrica especial, $x = x(t)$, $y = y(t)$, de un arco simple C tiene un *sentido* definido sobre C correspondiente al sentido de las *t crecientes*. Dados dos puntos distintos cualesquiera, P_0, P_1, se dice que P_1 sigue a P_0 si P_1 pertenece al mayor valor del parámetro t. Si se introduce un nuevo parámetro τ mediante una función continua *creciente* $\tau = \chi(t)$, el orden de los pares de puntos con respecto a τ es el mismo; y el parámetro τ define el *mismo sentido* sobre C. Si $\chi(t)$ es decreciente, el sentido se invierte.

Dirección u orientación de arcos

Un arco simple *dirigido* u *orientado* es aquél en el cual se ha escogido un sentido definido (por ejemplo, el sentido que corresponde a un incre-

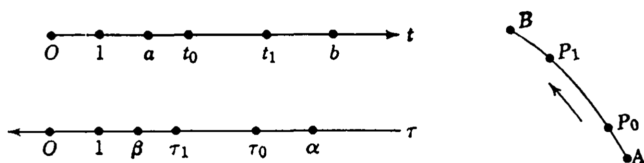


Figura 4.8 Sentido y representación paramétrica.

mento para una elección particular del parámetro t); y ese sentido se llama entonces el *sentido positivo* del arco. El sentido positivo queda completamente especificado si se sabe cuál de los dos puntos extremos del arco *sigue* al otro. Se denominará punto *final* del arco al punto extremo que sigue al otro, y punto *inicial* a éste. Dada cualquier representación paramétrica $x = x(\tau)$, $y = y(\tau)$ del arco orientado, donde $a \leq \tau \leq b$, el sentido positivo será aquél que corresponde al sentido de las τ crecientes si el valor paramétrico $\tau = a$ corresponde al punto inicial y el $\tau = b$, al punto final; en caso contrario, el sentido de las τ crecientes corresponderá al sentido negativo del arco (Fig. 4.8).

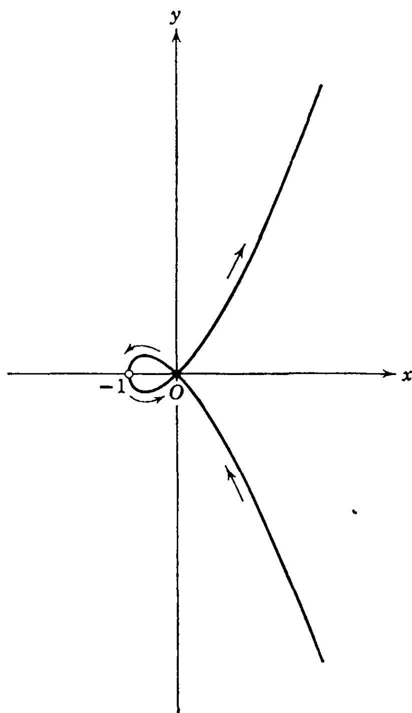


Figura 4.9 Una curva con una espira: $x = t^2 - 1$, $y = t^3 - t$, cuyo sentido es el de las t crecientes.

Dos puntos distintos cualesquiera, P_0, P_1, sobre un arco simple C definen un subarco con puntos extremos P_0, P_1, el cual consiste de los puntos con valores paramétricos comprendidos entre los correspondientes para P_0 y P_1. Si C es un arco dirigido y P_1 sigue a P_0 en el sentido positivo de C, se obtiene un subarco dirigido con punto inicial P_0 y punto final P_1. Un número finito de puntos de subdivisión sobre un arco simple dirigido C parte dicho arco en una sucesión de subarcos dirigidos, siendo el punto inicial de un subarco el punto final del que le precede.

A menudo resulta impráctico restringirse a arcos simples e insistir en que diferentes valores del parámetro t pertenezcan a puntos diferentes sobre la curva. Si, por ejemplo, las ecuaciones $x = x(t)$, $y = y(t)$ dan la posición de una partícula móvil P en el tiempo t, no existe razón para que la partícula no pueda permanecer en reposo durante algún tiempo o para que su trayectoria no se corte a sí misma de manera que la partícula regrese a la misma posición un tiempo después.

Un ejemplo lo constituye la curva $x = t^2 - 1$, $y = t^3 - t$ [la cual también podría describirse completamente mediante la ecuación cúbica $y^2 - x^2(1 + x) = 0$]. Conforme t varía desde $-\infty$ hasta $+\infty$, la curva pasa por el origen dos veces, para $t = -1$ y para $t = +1$ (Fig. 4.9). Fácilmente se verifica que los demás puntos de la curva pertenecen a un único valor de t. Geométricamente, el intervalo $-1 < t < +1$ corresponde a una *espira* de la curva. Aquí, nuevamente, el sentido de las t crecientes define un cierto *orden entre los puntos de la curva,* al menos si se interpretan de alguna manera los puntos correspondientes a $t = -1$ y a $t = +1$ como diferentes, estando uno de ellos situado "encima" del otro. La curva cúbica completa y orientada puede descomponerse en arcos simples dirigidos, por ejemplo, en los arcos correspondientes a $n \leq t \leq n + 1$, donde n recorre todos los enteros.

Curvas cerradas

El ejemplo usual de una representación paramétrica en la cual valores de t diferentes corresponden al mismo punto de la curva está dado por las fórmulas

$$x = a \cos t, \qquad y = a \operatorname{sen} t,$$

que describen el movimiento uniforme de un punto sobre una circunferencia, siendo t el tiempo. Conforme t varía desde $-\infty$ hasta $+\infty$ el punto $P = (x, y)$ recorre la circunferencia un número infinito de veces en el sentido contra reloj. Podemos hacer que los puntos de la circunferencia sean recorridos exactamente una vez restringiendo t a cualquier intervalo semiabierto de longitud 2π:

$$\alpha \leq t < \alpha + 2\pi.$$

Los puntos extremos α y $\alpha + 2\pi$ del intervalo corresponden al mismo punto de la circunferencia. Aquí los puntos extremos del intervalo paramétrico no tienen un significado geométrico especial para la curva.

Generalmente, un par de funciones continuas, $x = \phi(t)$, $y = \psi(t)$, definidas en un intervalo cerrado $a \leq t \leq b$ representan una curva *cerrada* si $\phi(a) = \phi(b)$ y $\psi(a) = \psi(b)$. La curva cerrada será simple si valores t diferentes, con $a \leq t < b$, corresponden a puntos diferentes (x, y).

El punto correspondiente a $t = a$ y $t = b$ podría ser cualquier punto de la curva; y es justamente el punto en el cual la curva se "rompe" para hacer que sus puntos correspondan a los de un intervalo sobre el eje.

Curvas cerradas representadas por funciones periódicas

Justamente como en el ejemplo de la circunferencia, se puede evitar el distinguir cualquier punto de división particular tomando para $\phi(t)$ y $\psi(t)$ funciones *periódicas* con período $p = b - a$. Es sumamente valioso hacer aquí algunos comentarios generales acerca de las funciones periódicas, a las cuales regresaremos para tratarlas en forma más intensiva en el capítulo 8.

Una función $f(t)$ se llama *periódica* con período p si está definida para todo t y satisface la ecuación $f(t) = f(t + p)$. Así, por ejemplo, las funciones trigonométricas sen t y cos t son periódicas con período 2π. (Cualquier múltiplo $2n\pi$, donde n es un entero, es entonces también un período.) Interpretada geométricamente, $f(t)$ tiene período p si un corrimiento de su gráfica en p unidades hacia la derecha lleva nuevamente a la misma gráfica.

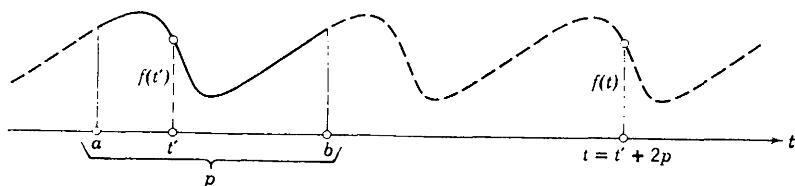


Figura 4.10 Gráfica de una función periódica. $f(t)$.

Puesto que $f(t)$ "se repite" a sí misma, una función $f(t)$ de período p está determinada para todo t si se conoce dicha función en un intervalo $a \leq t < b$ de longitud $p = b - a$ (Fig. 4.10). En efecto, para cada t existe un valor t' en el intervalo $a \leq t' < b$ tal que $t - t' = np$, donde n es un entero [sólo se tiene que tomar para n el mayor entero que no exceda a $(t - a)/p$]. Entonces $f(t) = f(t')$ es conocida.

De hecho, podemos partir con *cualquier* función $f(t)$ *continua* en un intervalo semiabierto $a \leq t < b$; y la función extendida será, claramente, continua para todo t que no sea de la forma $t = a + np$ con n entero (Fig. 4.11).

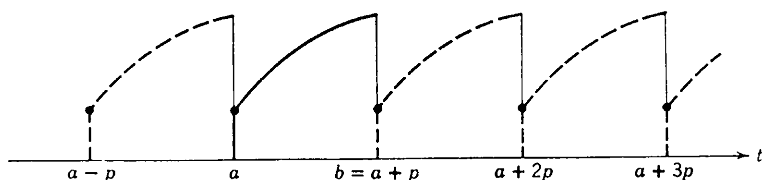


Figura 4.11 Continuación periódica de una función $f(t)$ partiendo del intervalo $a \leq t < b$.

Por ejemplo, la extensión en forma periódica de la función $f(t)$ definida por $f(t) = t$ para $0 \leq t < 1$, conduce a una función de período $p = 1$, que llamaremos "parte fraccional de t", y la cual es discontinua en los puntos t que son enteros (Fig. 4.12*a*). Generalmente, en $t = a + np$ la función periódicamente extendida tendrá el valor $f(a)$; y éste también será el límite de f al aproximarse al punto desde la derecha, en tanto que el límite de f desde la izquierda será el mismo que el correspondiente al punto b. En el caso de mayor interés para nosotros por ahora, comenzamos con una función definida y continua en el intervalo *cerrado* $a \leq t \leq b$, la cual tiene además el mismo valor en los puntos extremos: $f(a) = f(b)$. Continuando periódicamente tal función se llega siempre a una función $f(t)$ de período $p = b - a$, la cual es continua para todo t. (Fig. 4.12*b*).

Las funciones periódicas continuas son ideales para representar curvas cerradas C. Sea C dada paramétricamente por $x = \phi(t)$, $y = \psi(t)$, para

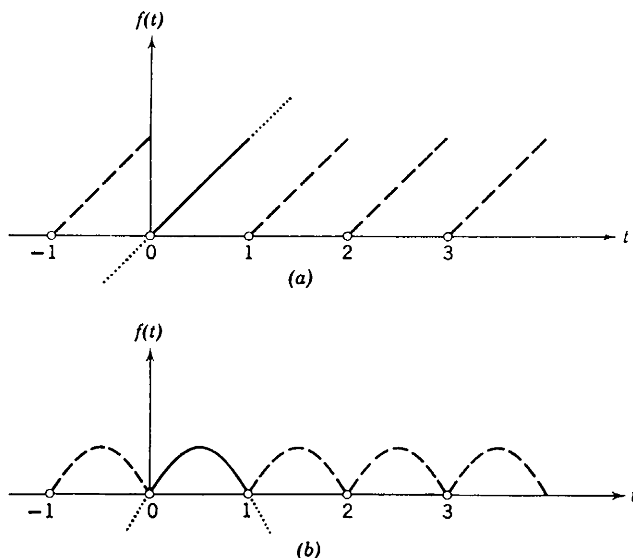


Figura 4.12 Continuación periódica de las funciones $f(t)$ definidas inicialmente en el intervalo $0 \leq t < 1$. Aquí (*a*) $f(t) = t$, (*b*) $f(t) = 2t - 2t^2$.

ϕ, ψ continuas en el intervalo $a \leq t \leq b$ y teniendo los mismos valores en ambos puntos extremos. Se puede extender la definición de estas funciones a todos los valores de t de manera que ϕ y ψ tengan período $b - a = p$ y sean continuas para todo t. Para cualquier t la representación paramétrica extendida solamente conduce a puntos de C puesto que se tiene $t = t' + np$ con n entero y $a \leq t' \leq b$. El punto correspondiente a t es entonces el mismo que el correspondiente a t', el cual está sobre C. Conforme t varía desde $-\infty$ hasta $+\infty$ el punto (x, y) recorre la curva C un número in-

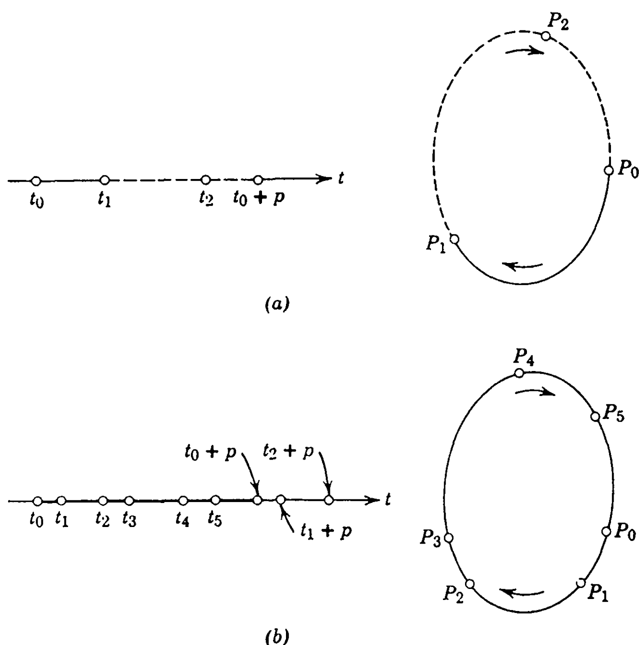


Figura 4.13 Orientación de curvas cerradas en el sentido de las t crecientes.

finito de veces, justamente como en el caso de la circunferencia $x = a \cos t$, $y = a \operatorname{sen} t$. Aquí el papel sobresaliente del valor paramétrico $t = a$ es eliminado. Para cualquier α la curva completa queda ahora representada por $x = \phi(t)$, $y = \psi(t)$ cuando t varía desde α hasta $\alpha + p$.

Una porción de la curva cerrada C correspondiente a los valores del parámetro t en un intervalo $\alpha \leq t \leq \beta$ forma un arco simple si valores t diferentes en ese intervalo conducen a puntos diferentes (x, y). La curva cerrada C completa es una curva simple si diferentes t en el mismo intervalo $\alpha \leq t < \alpha + p$ siempre conducen a puntos diferentes sobre C. Así, cualquier intervalo paramétrico cerrado de longitud menor que p origina un arco simple.

Curvas cerradas compuestas de arcos simples. Orden de puntos

Las curvas cerradas que se considerarán pueden todas ellas descomponerse en arcos simples. Si la curva cerrada C completa es simple, puede descomponerse en dos arcos simples $t_0 \leq t \leq t_1$ y $t_1 \leq t \leq t_0 + p$ que tengan solamente sus puntos extremos, P_0, P_1, en común. El sentido de las t crecientes determina un sentido positivo u orientación sobre C, fijando una dirección positiva sobre cada arco simple de C. Dos puntos distintos cualesquiera, P_0, P_1, sobre la curva cerrada simple C dividen a ésta en dos arcos simples. En el sentido de las t crecientes exactamente uno de los dos arcos tendrá a P_0 como punto inicial y a P_1 como punto final; lo llamaremos P_0P_1. La inversa se cumple para el otro arco.

Orientación y orden

La orientación positiva de C puede también caracterizarse mediante una terna ordenada de puntos $P_0P_1P_2$ de C, si se especifica que P_2 no está sobre el arco dirigido simple con punto inicial P_0 y punto final P_1. Las ternas $P_1P_2P_0$ y $P_2P_0P_1$ obtenidas por permutación cíclica a partir de $P_0P_1P_2$ describen la misma orientación (Fig. 4.13*a*).

*En forma completamente general, cualesquier n puntos distintos sobre la curva cerrada simple C, orientada, siempre se suceden uno al otro en un orden determinado $P_1P_2 \cdots P_n$, excepto por permutaciones cíclicas,[1] y dividen C en arcos simples dirigidos $P_1P_2, \ldots, P_{n-1}P_n$, P_nP_1. Pueden siempre escogerse valores paramétricos $t_1, t_2, \ldots, t_n$ para los puntos $P_1, P_2, \ldots, P_n$ de manera tal que los t_i formen una sucesión monótona creciente y estén todos contenidos en uno y el mismo intervalo paramétrico de longitud igual al período p (Fig. 4.13*b*).

Orientación de curvas y ángulos

Como ya se ha señalado en el capítulo 1, estamos obligados a hacer uso de los signos más y menos para establecer relaciones satisfactorias entre los entes geométricos y los conceptos analíticos expresados por números. Los ejemplos más simples los contituyen las *rectas dirigidas,* tales como el eje numérico. El sentido que se define como positivo en una recta es arbitrario al principio. Un sentido positivo correspondiente a valores crecientes de t puede asociarse con cualquier representación paramétrica particular $x = at + b$, $y = ct + d$ de la recta. Una recta orientada de esta manera *apunta* en una cierta *dirección.* Dos rectas paralelas dirigidas tienen el mismo sentido o bien sentidos opuestos. Una dirección puede también determinarse mediante un *rayo* que parte de un punto P_0, esto es, mediante una semirrecta que consiste de todos los puntos de la recta que "siguen" en el sentido positivo a un punto dado P_0.

[1] Esto es, $P_2P_3 \cdots P_nP_1$, $P_3P_4 \cdots P_nP_1P_2$, $\ldots$, $P_nP_1 \cdots P_{n-1}$ dan la misma orientación.

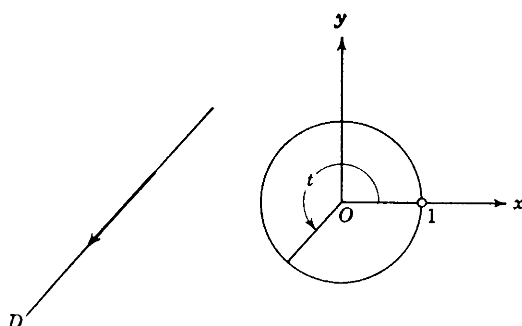


Figura 4.14 Angulo de inclinación t de una dirección D.

Cualquier dirección en el plano puede representarse mediante un rayo que parte del origen, y también mediante el punto P que es común al rayo y a la circunferencia de radio 1 con centro en el origen. Si representamos esta circunferencia unitaria en forma paramétrica mediante $x = \cos t$, $y = \operatorname{sen} t$, habremos asociado con cada dirección ciertos valores t que difieren uno de otro por múltiplos de 2π. Estos son los llamados *ángulos de inclinación* de la dirección o ángulos que la dirección forma con el eje x positivo. Existe siempre exactamente un ángulo de inclinación t para el cual $0 \leq t < 2\pi$ (Fig. 4.14).

Los *ángulos entre dos direcciones* son simplemente las diferencias entre sus ángulos de inclinación. Más precisamente, puesto que el orden en el cual se toman las dos direcciones es importante, decimos que *una dirección de inclinación t' forma con una dirección de inclinación t'' un ángulo* $\alpha = t' - t''$ (Fig. 4.15). Como t y t' pueden cambiar en múltiplos enteros de 2π, el mismo cambio es permisible para el ángulo que una dirección forma con otra.

Sentido de rotación

También se dice que la dirección con ángulo de inclinación t'' se convierte en la dirección t' mediante una *rotación* a través de un ángulo α. La

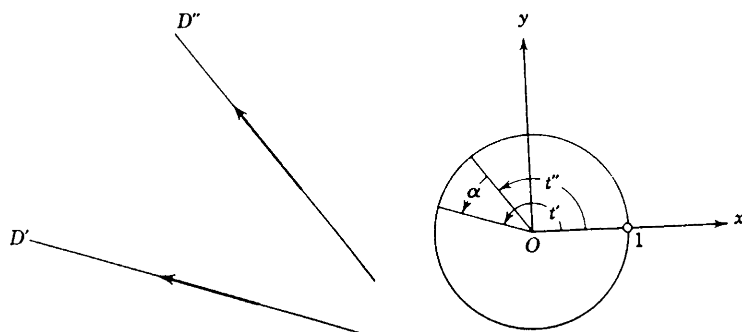


Figura 4.15 Angulo α que la dirección D' forma con la dirección D''.

idea intuitiva de rotación es aquí la de un *movimiento continuo* mediante el cual la dirección con inclinación t'' se transforma en la dirección con inclinación t' pasando por las direcciones con todas las inclinaciones posibles t intermedias entre t'' y t'. La rotación se denomina *positiva* o *contra reloj* si $\alpha = t' - t''$ es positivo, y *negativa* o *en el sentido del reloj* para el caso contrario. Por supuesto, existen muchas rotaciones diferentes, tanto en el sentido del reloj como en el opuesto, que convertirán una cierta dirección en otra dirección dada, a menos que se especifique que el ángulo de rotación satisface $-\pi < \alpha \leq \pi$.

Entonces, finalmente, el sentido positivo de rotación está asociado con una representación paramétrica particular $x = \cos t$, $y = \operatorname{sen} t$ de la circunferencia que se ha escogido. Si, como es costumbre, el eje x apunta hacia la derecha y el eje y hacia arriba, entonces el sentido positivo de rotación coincide con el sentido opuesto al de las manecillas de un reloj convencional.[1]

Lados positivo y negativo de una curva

Una curva separa los puntos del plano cercanos a uno de sus puntos P en dos clases. Por lo menos localmente pueden distinguirse dos "lados" de la curva. Si la curva C está orientada podemos definir un lado *positivo* (o "izquierdo")[2] y un lado *negativo* (o "derecho") de la siguiente manera. Considérese un rayo que parte de P. Se dice que este rayo apunta hacia el lado positivo de una curva si existen puntos Q sobre la curva, arbitrariamente próximos a P y que siguen a P en el sentido dado a la curva, tales que la recta PQ deba girarse en el sentido contra reloj un ángulo comprendido entre 0 y π para llegar al rayo dado (Fig. 4.16). Los puntos sobre el rayo cercanos a P se dice que están sobre el lado positivo de la curva. En el caso opuesto se dice que el rayo apunta hacia el lado negativo de C y que los puntos sobre él están en el lado negativo de la curva. Si C es una curva cerrada simple, ella divide todos los puntos del plano en dos clases, los *interiores* a C y los *exteriores* a C.[3] Se dice que C tiene una *orientación*

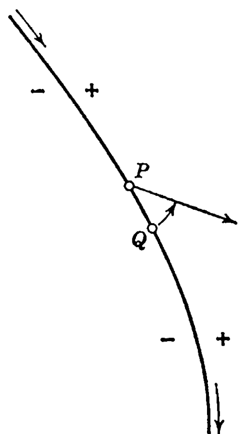


Figura 4.16 Lados positivo y negativo de un arco orientado.

[1] Este sentido, a su vez, es sugerido por el movimiento en el suelo de la sombra de un reloj de Sol en el hemisferio norte.

[2] Los términos "lado derecho" e "izquierdo" corresponden al uso ordinario de las palabras "ribera izquierda" y "ribera derecha" para un río orientado según la dirección de su flujo.

[3] Estos conceptos, así como la división del plano en dos partes por medio de una curva continua, cerrada, simple, se analizan en forma precisa en la topología. Aquí deben aceptarse sobre una base intuitiva.

contra reloj si su interior está sobre el lado positivo (esto es, izquierdo) (Fig. 4.17).

Pero si la curva cerrada C consiste de varias espiras, entonces no siempre es posible recorrer la curva de manera que todas las regiones encerradas estén en el lado positivo de ella (véase Fig. 4.18).

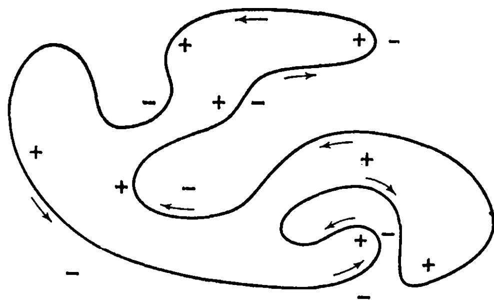


Figura 4.17 Curva cerrada simple con orientación contra reloj.

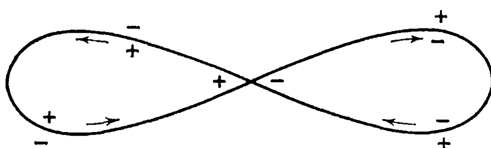


Figura 4.18

e. Derivadas, tangentes y normales en representación paramétrica

Dirección y velocidad

Para una curva C dada en representación paramétrica usando el tiempo como parámetro t,

$$x = x(t) = \phi(t), \quad y = y(t) = \psi(t),$$

las derivadas se denotan, como lo hizo Newton, mediante un punto:

$$\dot{x} = \frac{d\phi}{dt} = \dot{\phi}, \quad \dot{y} = \frac{d\psi}{dt} = \dot{\psi}.$$

Las derivadas $\dot{x}$ e $\dot{y}$ son a menudo visualizadas convenientemente como las "componentes de velocidad" o las "velocidades" de las coordenadas de un punto P que se mueve a lo largo de C.

Siempre que $\dot{x} \neq 0$, será posible representar la correspondiente porción de C por medio de una ecuación $y = f(x)$ calculando primero t como una función de x a partir de la primera ecuación y luego substituyendo la expresión resultante para t en la segunda ecuación. Por la regla de la cadena

de la derivación y la regla para la derivada de la inversa de una función (ver p. 228), se encuentra que la pendiente de la tangente a la curva está dada por

$$\frac{dy}{dx} = \frac{dy}{dt}\,\frac{dt}{dx} = \frac{\dfrac{dy}{dt}}{\dfrac{dx}{dt}} = \frac{\dot{y}}{\dot{x}}.$$

La fórmula equivalente $dx/dy = \dot{x}/\dot{y}$ se cumple si $\dot{y} \neq 0$.

A menos que se especifique lo contrario, siempre se supondrá que $\dot{x}$ e $\dot{y}$ no se anulan simultáneamente, o sea, escrito en forma concisa, se supondrá que

$$\dot{x}^2 + \dot{y}^2 \neq 0.$$

Entonces la tangente siempre existe;[1] es horizontal si $\dot{y} = 0$ y vertical si $\dot{x} = 0$.

Para la *cicloide,* por ejemplo [ver Ec. (1), p. 347] se tiene

$$\dot{x} = a(1 - \cos t) = 2a \operatorname{sen}^2 \frac{t}{2},$$

$$\dot{y} = a \operatorname{sen} t = 2a \operatorname{sen} \frac{t}{2} \cos \frac{t}{2},$$

$$\frac{dy}{dx} = \cot \frac{t}{2}.$$

Estas fórmulas muestran que $\dot{x}^2 + \dot{y}^2 \neq 0$ excepto para $t = 0, \pm 2\pi, \pm 4\pi, \ldots$. Además, la cicloide tiene una cúspide (esto es, un punto en el que se invierte la dirección) con una tangente vertical en esos puntos excepcionales, en los que también encuentra al eje x, esto es, para los cuales $y = 0$. En efecto, al aproximarse a estos puntos la derivada $y' = \dot{y}/\dot{x} = \cot(t/2)$ se vuelve infinita.

Tangente, normal y cosenos directores

La ecuación de la tangente a la curva en el punto x, y es

$$\eta - y = \frac{dy}{dx}(\xi - x),$$

donde ξ y η son las coordenadas "variables" correspondientes a un punto arbitrario sobre la tangente, en tanto que x, y y dy/dx tienen los valores

[1] Observamos que la condición $\dot{x}^2 + \dot{y}^2 \neq 0$, aunque suficiente, no es necesaria para garantizar una representación no paramétrica. Así, podemos definir la curva $y = x^2$ por medio de las ecuaciones paramétricas $x = t^3$, $y = t^6$. En el origen del eje t la condición de positividad para $\dot{x}^2 + \dot{y}^2$ falla, pero aún la curva tiene una representación no paramétrica bien definida.

fijos que corresponden al punto de contacto. Substituyendo dy/dx por $\dot{y}/\dot{x}$ podemos escribir la ecuación de la tangente en la forma

$$(5) \qquad (\xi - x)\dot{y} - (\eta - y)\dot{x} = 0.$$

La misma ecuación exactamente se obtiene con la hipótesis $\dot{y} \neq 0$; y lo único que se necesita es expresar x como función de y. En los puntos excepcionales para los que tanto $\dot{x}$ como $\dot{y}$ se anulan para el mismo valor t, la ecuación pierde todo significado, pues entonces es satisfecha por ξ, η cualesquiera.

La *normal a la curva,* esto es, la recta que pasa por un punto de la curva y es perpendicular a la tangente en ese punto, tiene la pendiente $-dx/dy$. Esto conduce a la ecuación

$$(6) \qquad (\xi - x)\dot{x} + (\eta - y)\dot{y} = 0$$

para la normal.

Si un punto de C corresponde a varios valores de t, entonces en general existe una tangente diferente para cada una de las ramas de la curva que pasan por el punto, o para cada valor de t. Por ejemplo, la curva $x = t^2 - 1$, $y = t^3 - t$ (Fig. 4.9, p. 353) pasa por el origen para $t = -1$ y $t = +1$. Para $t = -1$ se encuentra que la ecuación de la tangente es $\xi + \eta = 0$, en tanto que para $t = +1$ la tangente está dada por $\xi - \eta = 0$.

De la definición de derivada se tiene

$$\frac{dy}{dx} = \frac{\dot{y}}{\dot{x}} = \tan\alpha,$$

donde α es el ángulo que la tangente forma con el eje x. Esto significa que una rotación del eje x a través del ángulo α (contra reloj si $\alpha > 0$, en el sentido del reloj si $\alpha < 0$) hará que dicho eje sea paralelo a la tangente. Entonces también rotaciones a través de ángulos $\alpha \pm \pi$, $\alpha \pm 2\pi, \ldots$ harán el eje x paralelo a dicha tangente. Por lo tanto, el ángulo α está determinado solamente salvo un múltiplo de π, en tanto que $\tan\alpha$ está determinada en forma única. De las relaciones $\dot{y}/\dot{x} = (\operatorname{sen}\alpha)/(\cos\alpha)$ y $\dot{x}^2 + \dot{y}^2 \neq 0$ se obtiene

$$\cos\alpha = \pm\frac{\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2}}, \qquad \operatorname{sen}\alpha = \pm\frac{\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}}$$

(debiendo usarse el mismo signo en ambas fórmulas). A $\cos\alpha$ y $\operatorname{sen}\alpha$ se les llama los *cosenos directores* de la tangente.[1]

Asignación de direcciones a la tangente y a la normal

Las dos posibles elecciones para los cosenos directores corresponden a las dos direcciones en las cuales puede recorrerse la tangente; y los corres-

[1] Aquí se concibe $\operatorname{sen}\alpha$ como $\cos\beta$, donde $\beta = \pi/2 - \alpha$ es el ángulo que el eje y forma con la tangente.

pondientes ángulos α difieren entre sí en un múltiplo impar de π. Una de las dos direcciones de la tangente corresponde a t creciente; la otra, a t decreciente. Supóngase que la dirección de la curva es la asociada a las t crecientes. Entonces, por definición, la dirección positiva de la tangente, o la que corresponde a valores crecientes de t, es aquélla que forma con el eje x positivo un ángulo α para el cual $\cos \alpha$ tiene el mismo signo que $\dot{x}$, y $\operatorname{sen} \alpha$ el mismo signo que $\dot{y}$. Los cosenos directores correspondientes a esa dirección de la tangente son entonces, sin ambigüedad:

$$(7) \qquad \cos \alpha = \frac{\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2}}, \qquad \operatorname{sen} \alpha = \frac{\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}}.$$

Si, por ejemplo, $\dot{x} = dx/dt > 0$, la dirección de la tangente que corresponde a las t crecientes es también la que corresponde a las x crecientes;

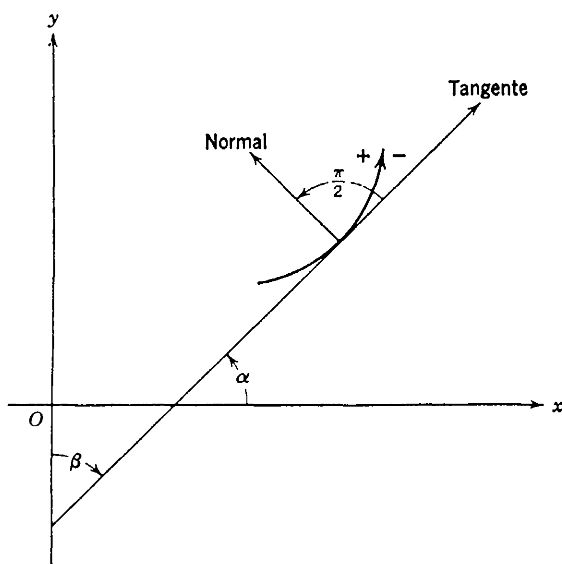


Figura 4.19 Tangente y normal positiva de una curva orientada.

y el ángulo que esa dirección forma con el eje x positivo tiene entonces un coseno positivo. Análogamente, la dirección normal que se obtiene haciendo girar un ángulo $\pi/2$ en el sentido positivo (contra reloj) la dirección de la tangente positiva, que corresponde a valores crecientes de t, tiene como cosenos directores

$$\cos\left(\alpha + \frac{\pi}{2}\right) = \frac{-\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}}, \qquad \operatorname{sen}\left(\alpha + \frac{\pi}{2}\right) = \frac{\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2}}$$

Esta es la llamada dirección normal positiva y apunta hacia el "lado positivo" de la curva (Fig. 4.19).

Si se introduce un nuevo parámetro $\tau = \chi(t)$ para la curva, los valores de $\cos\alpha$ y $\operatorname{sen}\alpha$ no cambian si $d\tau/dt > 0$ y cambian de signo si $d\tau/dt < 0$; esto es, si cambia el sentido de la curva cambian también los sentidos positivos de la tangente y de la normal.

Puntos críticos

Si $\dot{x}$ e $\dot{y}$ son continuas y $\dot{x}^2 + \dot{y}^2 > 0$, las cantidades $\cos\alpha$ y $\operatorname{sen}\alpha$ que determinan la dirección de la tangente variarán continuamente con t. La tangente, cuya ecuación es

$$(\xi - x)\operatorname{sen}\alpha - (\eta - y)\cos\alpha = 0,$$

cambia entonces en forma continua a lo largo de la curva; lo mismo que la normal.

Si $\dot{x}$ e $\dot{y}$ se anulan para cierto valor de t, los cosenos directores de la tangente no quedan definidos por nuestras fórmulas; y la tangente puede no existir en absoluto o no estar definida en forma única. A tales puntos se les llama puntos "críticos" o puntos "estacionarios". Mediante ejemplos se ilustrarán varias posibilidades en relación con los puntos críticos.

Uno de tales ejemplos lo proporciona la curva $y = |x|$ con la representación paramétrica $x = t^3$, $y = |t|^3$; esta curva tiene una esquina para $t = 0$, aunque tanto $\dot{x}$ como $\dot{y}$ son continuas. En el ejemplo de la cicloide, discutido en la p. 362, los puntos "estacionarios" en los cuales $\dot{x} = \dot{y} = 0$ corresponden a cúspides. Por otra parte, en algunos casos la anulación de $\dot{x}$ y de $\dot{y}$ es sólo consecuencia de la representación paramétrica específica y no tiene relación con el comportamiento de la curva, como ocurre para la recta representada por $x = t^3$, $y = t^3$, para el valor paramétrico $t = 0$.

Esquinas

Las curvas que consisten de varios arcos suaves que se unen en esquinas son representadas convenientemente en forma paramétrica mediante

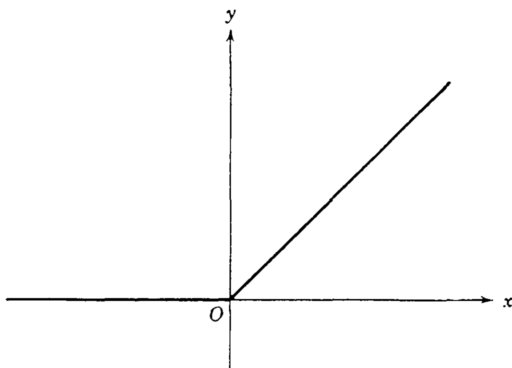


Figura 4.20 Gráfica de $x = t$, $y = \frac{1}{2}(t + |t|)$.

funciones $x(t)$, $y(t)$ continuas pero con derivadas $\dot{x}$, $\dot{y}$ que presentan discontinuidades de salto. Esto queda ilustrado por el ejemplo trivial de la línea quebrada con representación

$$x = t, \qquad y = 0 \qquad \text{para} \quad t \leq 0$$

y

$$x = t, \qquad y = t \qquad \text{para} \quad t \geq 0.$$

Aquí $\dot{x} = 1$, $\dot{y} = 0$ para $t < 0$, y $\dot{x} = 1$, $\dot{y} = 1$ para $t > 0$. En $t = 0$ la tangente está indeterminada (ver Fig. 4.20).

f. La longitud de una curva

La longitud como una integral

Existen dos tipos diferentes de cantidades o propiedades geométricas asociadas con una curva. El primer tipo depende sólo del comportamiento *local* de la curva, esto es, en la inmediata vecindad de un punto; tales propiedades son aquéllas que pueden representarse mediante derivadas en el punto en cuestión. Propiedades del segundo tipo, las ***propiedades globales***, dependen de la configuración total de la curva o de una porción de ella y son usualmente expresadas en forma analítica mediante el concepto de integral. Empezaremos considerando una cantidad del segundo tipo, la longitud de una curva.

Por supuesto, se tiene ya una noción intuitiva de lo que se entiende por longitud de una curva. Sin embargo, como en el caso clásico de los arcos circulares, debe darse un significado matemático preciso al concepto intuitivo. Guiados por la intuición definimos la longitud de una curva arbitraria como el límite de las longitudes de polígonos que se le aproximan, en particular, polígonos inscritos. Las longitudes de los polígonos, a su vez, son definidas no bien se escoja una unidad de longitud. El resultado final será la expresión de la longitud mediante una integral.

Considérese la curva dada en la forma $x = x(t)$, $y = y(t)$, $\alpha \leq t \leq \beta$. En el intervalo comprendido entre α y β se escogen puntos intermedios $t_1, t_2, \ldots, t_{n-1}$ tales que

$$\alpha = t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = \beta.$$

Los puntos $P_0, P_1, \ldots, P_n$ de la curva que corresponden a estos valores t_i en el mismo orden se unen mediante segmentos de recta, obteniéndose así un polígono inscrito. La longitud del perímetro de este polígono inscrito depende de la manera en que los puntos t_i o los vértices P_i del polígono sean escogidos. Déjese ahora que el número de puntos t_i crezca más allá de toda cota, de manera que, al mismo tiempo, la longitud del subintervalo (t_i, t_{i+1}) más largo tienda a cero. La longitud de la curva se define entonces

como el límite de los perímetros de estos polígonos inscritos, *siempre que* tal límite exista y sea independiente de la manera particular como se escojan los polígonos. Cuando esta hipótesis (hipótesis de *rectificabilidad)* se satisface, podemos hablar de la longitud de la curva.

Se supondrá que las funciones $x(t)$ e $y(t)$ tienen derivadas $\dot{x}(t)$ e $\dot{y}(t)$ continuas para $\alpha \leq t \leq \beta$. El polígono inscrito correspondiente a la subdivisión del intervalo t mediante los puntos t_i, con $\Delta t_i = t_{i+1} - t_i$, tiene vértices $P_i = (x(t_i), y(t_i))$; su longitud total está dada por la expresión

$$S = \sum_{i=0}^{n-1} \overline{P_i P_{i+1}} = \sum_{i=0}^{n-1} \sqrt{[x(t_{i+1}) - x(t_i)]^2 + [y(t_{i+1}) - y(t_i)]^2},$$

de acuerdo con el teorema de Pitágoras (cfr. Fig. 4.21, p. 374). Por el teorema del valor medio del cálculo diferencial,

$$x(t_{i+1}) - x(t_i) = \dot{x}(\xi_i)\,\Delta t_i, \qquad y(t_{i+1}) - y(t_i) = \dot{y}(\eta_i)\,\Delta t_i,$$

donde ξ_i y η_i son valores intermedios en el intervalo $t_i < t < t_{i+1}$. Esto conduce a la expresión

$$S_n = \sum_{i=0}^{n-1} \sqrt{[\dot{x}^2(\xi_i)]^2 + [\dot{y}^2(\eta_i)]^2}\,\Delta t_i$$

para la longitud del polígono, donde se ha usado el hecho de que las diferencias Δt_i son positivas. Si el número n de puntos t_i de subdivisión crece más allá de toda cota en tanto que, simultáneamente, el máximo valor Δt_i tiende a cero, la suma S_n tiende a la integral

$$L = \int_\alpha^\beta \sqrt{\dot{x}^2 + \dot{y}^2\,dt.}$$

Este hceho es una consecuencia directa de los teoremas de existencia para integrales, del capítulo 2.[1]

Lo anterior prueba que, para $\dot{x}$, $\dot{y}$ continuas, la curva realmente tiene una longitud y que esta longitud está dada por la expresión

$$(8) \qquad L = \int_\alpha^\beta \sqrt{\dot{x}^2 + \dot{y}^2\,dt.}$$

La misma regla es válida si se admite que $\dot{x}$ e $\dot{y}$ sean discontinuas en puntos aislados, en los que la curva podría no tener, entonces, una tangente única; y, por supuesto, la integral debe entonces ser considerada como "impropia" (ver capítulo 3, p. 320). Curvas "rectificables" más generales, para las cuales la integral anterior sigue teniendo significado, no serán discutidas en este volumen.

---

[1] Puesto que los puntos intermedios ξ_i y η_i no necesariamente coinciden, hacemos uso de las sumas de aproximación más generales cuya convergencia a la integral se demostró en la p. 216.

Definición alternativa de longitud

Hacemos una observación interesante: El perímetro S de cualquier polígono inscrito π nunca puede exceder la longitud L de la curva. (En particular, la distancia entre los puntos extremos de la curva no puede exceder a L, pues la línea recta que une los puntos extremos es la curva *más corta* que va de uno de tales puntos al otro.) En efecto, L puede obtenerse como el límite de los perímetros de una sucesión especial de polígonos inscritos, en la cual se parte con el polígono π de perímetro S y se obtienen los siguientes agregando en forma sucesiva más y más vértices. Insertando un vértice adicional entre dos vértices sucesivos de un polígono inscrito nunca puede llegarse a una disminución de los perímetros, pues un lado de un triángulo jamás excede la suma de los otros dos. Así, L es el límite de una sucesión no decreciente de perímetros que comienza con S. Por tanto, $S \leq L$. Por consiguiente, en vez de definir L como el límite de los perímetros de una sucesión de polígonos inscritos correspondientes a subdivisiones cada vez más finas del intervalo t, podría también haberse definido L como el *extremo superior* [1] de los perímetros de *todos* los polígonos inscriptos. Es interesante el hecho de que la longitud pueda definirse sin recurrir formalmente a ningún paso al límite.

Invariancia de la longitud respecto de cambios de parámetros

Resulta claro de la definición que la longitud L de una curva C no puede depender de la representación paramétrica particular usada para C. Por lo tanto, si se introduce un nuevo parámetro $\tau = \chi(t)$, donde $d\tau/dt > 0$, nuestra fórmula integral para L debe dar el mismo valor independiente de que el parámetro usado sea t o τ. Esto puede verificarse inmediatamente a partir de la regla de la cadena de la derivación y de las leyes de substitución para integrales. En efecto, se tiene

$$\sqrt{\dot{x}^2+\dot{y}^2} = \sqrt{\left(\frac{dx}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2} = \sqrt{\left(\frac{dx}{d\tau}\frac{d\tau}{dt}\right)^2+\left(\frac{dy}{d\tau}\frac{d\tau}{dt}\right)^2}$$
$$= \sqrt{\left(\frac{dx}{d\tau}\right)^2+\left(\frac{dy}{d\tau}\right)^2}\,\frac{d\tau}{dt}\,;$$

por lo tanto, si $\chi(\alpha) = a$, $\chi(\beta) = b$,

$$L = \int_\alpha^\beta \sqrt{\dot{x}^2+\dot{y}^2}\,dt = \int_\alpha^\beta \sqrt{\left(\frac{dx}{d\tau}\right)^2+\left(\frac{dy}{d\tau}\right)^2}\,\frac{d\tau}{dt}\,dt$$
$$= \int_a^b \sqrt{\left(\frac{dx}{d\tau}\right)^2+\left(\frac{dy}{d\tau}\right)^2}\,d\tau,$$

de manera que la expresión de la longitud basada en el parámetro τ conduce al mismo valor de L. Si, por otra parte, $d\tau/dt < 0$, análogamente se encuentra

---

[1] Definido como *la menor de las cotas superiores,* o *supremum,* y denotado frecuentemente por los símbolos "*extr*" y "*sup*". (Cfr. pp. 120-121). *(N. del R.)*

$$\int_{\alpha}^{\beta}\sqrt{\dot{x}^2+\dot{y}^2}\,dt = -\int_{a}^{b}\sqrt{\left(\frac{dx}{d\tau}\right)^2+\left(\frac{dy}{d\tau}\right)^2}\,d\tau$$
$$=\int_{b}^{a}\sqrt{\left(\frac{dx}{d\tau}\right)^2+\left(\frac{dy}{d\tau}\right)^2}\,d\tau;$$

y el miembro derecho es nuevamente la integral correcta para la longitud de C referida al parámetro τ, puesto que ahora $b < a$ debido a que $\chi(t)$ es una función decreciente.

Para una curva dada en forma no paramétrica mediante la función $y = f(x)$, $a \leq x \leq b$, podemos introducir x como parámetro t. Entonces $\dot{x} = 1$, $\dot{y} = dy/dx$. La longitud de la curva queda entonces dada por

$$L = \int_{a}^{b}\sqrt{1+\left(\frac{dy}{dx}\right)^2}\,dx. \tag{9}$$

Ejemplos. Como un ejemplo se encuentra la longitud de un segmento de la parábola $y = \frac{1}{2}x^2$ correspondiente al intervalo $a \leq x \leq b$:

$$L = \int_{t}^{b}\sqrt{1+x^2}\,dx.$$

En este caso la substitución $x = \operatorname{senh} t$ (véase capítulo 3, p. 293) conduce a

$$\int_{ar\,\operatorname{senh} a}^{ar\,\operatorname{senh} b}\cosh^2 t\,dt = \tfrac{1}{2}\int_{ar\,\operatorname{senh} a}^{ar\,\operatorname{senh} b}(1+\cosh 2t)\,dt$$
$$= \tfrac{1}{2}(t+\operatorname{senh} t\cosh t)\Big|_{ar\,\operatorname{senh} a}^{ar\,\operatorname{senh} b}$$
$$= \tfrac{1}{2}(ar\,\operatorname{senh} b + b\sqrt{1+b^2} - ar\,\operatorname{senh} a\sqrt{1+a^2}).$$

Para una curva dada por la ecuación $r = r(\theta)$, $\alpha \leq \theta \leq \beta$, en coordenadas polares, se tiene la representación $x = r(\theta)\cos\theta$, $y = r(\theta)\operatorname{sen}\theta$. Escogiendo θ como parámetro se tiene

$$\dot{x} = \dot{r}\cos\theta - r\operatorname{sen}\theta, \qquad \dot{y} = \dot{r}\operatorname{sen}\theta + r\cos\theta, \qquad \dot{x}^2+\dot{y}^2 = r^2+\dot{r}^2.$$

Esto conduce a la expresión

$$L = \int_{\alpha}^{\beta}\sqrt{r^2+\left(\frac{dr}{d\theta}\right)^2}\,d\theta \tag{10}$$

para la longitud de una curva en coordenadas polares. Por ejemplo, para la circunferencia de radio a alrededor del origen se tiene la ecuación $r =$ constante $= a$, $0 \leq \theta \leq 2\pi$. Esto da para la longitud total de la circunferencia

$$L = \int_{0}^{2\pi} a\,d\theta = 2\pi a.$$

Aditividad de la longitud

Sea C una curva dada por $x = x(t)$, $y = y(t)$, $\alpha \leq t \leq \beta$, donde $\dot{x}$ e $\dot{y}$ son continuas. Sea γ cualquier valor entre α y β. Por las reglas generales para integrales se tiene

$$\int_\alpha^\beta \sqrt{\dot{x}^2 + \dot{y}^2}\,dt = \int_\alpha^\gamma \sqrt{\dot{x}^2 + \dot{y}^2}\,dt + \int_\gamma^\beta \sqrt{\dot{x}^2 + \dot{y}^2}\,dt.$$

Las integrales de la derecha representan respectivamente las longitudes de las porciones en las que C es dividida por el punto correspondiente a $t = \gamma$. Por lo tanto, la longitud de la curva completa es igual a la suma de las longitudes de sus partes.

No es necesario que $\dot{x}$ e $\dot{y}$ sean continuas. Las integrales existen también cuando $\dot{x}$ e $\dot{y}$ tienen un número finito de discontinuidades de salto, como ocurriría para una curva con esquinas. La longitud total de la curva es entonces la suma de las longitudes de las porciones suaves comprendidas entre esquinas. Comportamientos aún más singulares de $\dot{x}$ e $\dot{y}$ son permitidos siempre y cuando la expresión de la longitud tenga sentido como integral impropia.

g. La longitud de arco como parámetro

Hemos visto que una y la misma curva admite muchas representaciones paramétricas diferentes $x = x(t)$, $y = y(t)$. Cualquier función monótona de t puede usarse como parámetro en lugar de t. Sin embargo, para muchos propósitos es ventajoso tomar como referencia para C algún "parámetro usual" que de alguna manera sea distinguido geométricamente. La abscisa x o el ángulo polar θ no son adecuados para ese propósito si las curvas han de describirse globalmente; además, dichos parámetros dependen de la elección del sistema de coordenadas. La posibilidad de medir longitudes a lo largo de una curva proporciona un parámetro natural definido geométricamente al cual pueden ser referidos los puntos P de una curva rectificable. Este parámetro es la longitud de la porción de la curva comprendida entre P y algún punto fijo P_0.

Comenzamos con una representación paramétrica arbitraria $x = x(t)$, $y = y(t)$, $\alpha \leq t \leq \beta$ de C. La derivación con respecto a t es indicada mediante un punto. Introducimos la "longitud de arco" s por medio de la integral indefinida

$$s = \int \sqrt{\dot{x}^2 + \dot{y}^2}\,dt, \tag{11}$$

o, en forma más precisa, s como una función de t por

$$s = s(t) = c + \int_{t_0}^{t} \sqrt{\dot{x}^2(\tau) + \dot{y}^2(\tau)}\,d\tau, \tag{11a}$$

donde c es una constante, t_0 es un valor entre α y β, y con τ se designa la variable de integración para distinguirla del límite superior t. Claramente, para cualesquier valores t_1 y t_2 del intervalo paramétrico la diferencia

$$(12) \qquad s(t_2) - s(t_1) = \int_{t_1}^{t_2} \sqrt{\dot{x}^2 + \dot{y}^2}\, d\tau$$

es igual a la longitud del pedazo de curva limitado por los puntos correspondientes a $t = t_1$ y $t = t_2$, suponiéndose $t_1 < t_2$. Para $t_1 > t_2$ la diferencia $s(t_2) - s(t_1)$ es la negativa de la longitud de esa porción de curva. Así, el conocimiento de cualquier integral indefinida s permite calcular la longitud de cualquier parte de la curva.

El signo de la longitud de arco

Si la constante c tiene el valor 0, puede interpretarse $s(t)$ como la longitud de arco de la curva (o "distancia a lo largo de la curva") entre el punto P_0 con parámetro t_0 y el punto P con parámetro t. Aquí la longitud se cuenta positiva en el caso en que el arco con punto inicial P_0 y punto final P tenga la orientación correspondiente a las t crecientes.[1]

La forma integral de la definición de s es equivalente a la relación

$$(12a) \qquad \frac{ds}{dt} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}.$$

Usando la notación simbólica para diferenciales (p. 201), $ds = (ds/dt)dt$, etc., podemos escribir esa relación en la sugestiva forma

$$ds = \sqrt{dx^2 + dy^2}$$

para el "elemento de longitud" ds.

Velocidad de movimiento a lo largo de una curva

Si t se interpreta como el tiempo y $x(t)$, $y(t)$ como las coordenadas de posición en el tiempo t de un punto o partícula puntual móvil, se tiene que

$$\dot{s} = \frac{ds}{dt} = \lim_{h \to 0} \frac{s(t+h) - s(t)}{h}$$

da la medida de la variación respecto al tiempo de la distancia recorrida por el punto *a lo largo de su trayectoria*, esto es, la *velocidad* de la partícula en el instante t.[2] Para una partícula que se mueve con velocidad

---

[1] Nótese que la variable s no es única; depende de la elección de P_0 y c y también de la orientación inducida por el parámetro t sobre la curva. Sin embargo, cualquier otra longitud de arco es expresable en términos de s en la forma $(s +$ constante$)$ o bien $(-s +$ constante$)$.

[2] O sea, el límite de la razón incremental $\Delta s/h$ de espacio sobre tiempo h. El cociente $\Delta s/h$ se llama *velocidad media* del punto en el intervalo $(t, t+h)$. *(N. del R.)*

uniforme a lo largo de la curva $\dot{s}$ es constante y s es entonces una función lineal del tiempo t.

Si nuestra hipótesis usual de que

$$\dot{x}^2 + \dot{y}^2 \neq 0$$

se satisface, tenemos $ds/dt \neq 0$ y podemos entonces introducir s como parámetro. De esta manera se simplifican una gran cantidad de fórmulas y cálculos. Las cantidades

$$\frac{dx}{ds} = \frac{dx}{dt}\frac{dt}{ds} = \frac{\dot{x}}{\sqrt{\dot{x}^2 - \dot{y}^2}} \qquad \frac{dy}{ds} = \frac{dy}{dt}\frac{dt}{ds} = \frac{\dot{y}}{\sqrt{\dot{x}^2 - \dot{y}^2}}$$

son entonces justamente los cosenos directores de la tangente que apunta en la dirección de s creciente (ver (7), p. 364). La relación

$$(13) \qquad \left(\frac{dx}{ds}\right)^2 + \left(\frac{dy}{ds}\right)^2 = 1$$

caracteriza el parámetro s como la longitud de arco a lo largo de la curva.

h. Curvatura

Definición mediante la medida del cambio de dirección

Se discute a continuación un concepto básico que se refiere solamente al *comportamiento local* de una curva en la vecindad de un punto, el concepto de curvatura.

Conforme se recorre la curva, el ángulo de inclinación α de la tangente a la curva variará en una medida definida por unidad de longitud de arco recorrida. Llamaremos a esta medida del cambio de α *la curvatura de la curva*. De acuerdo con esto, la curvatura está definida como

$$(14) \qquad \kappa = \frac{d\alpha}{ds}.$$

Expresiones paramétricas. Supóngase que la curva está dada paramétricamente mediante las funciones $x = x(t)$, $y = y(t)$ que tienen primera y segunda derivadas con respecto a t continuas, para las cuales se cumple $\dot{x}^2 + \dot{y}^2 \neq 0$. Al calcular $d\alpha/ds$, o sea, la medida de cambio del ángulo de dirección α en el punto P, se tiene que tomar en cuenta que α no está definido en forma única. Sin embargo, la función trigonométrica de α, $\tan\alpha = \dot{y}/\dot{x}$ (o bien $\cot\alpha = \dot{x}/\dot{y}$ para $\dot{x} = 0$) tiene un valor definido. Para formar $d\alpha/ds$ puede siempre suponerse que los valores paramétricos pertenecientes a puntos en la vecindad de P están todos ellos en un intervalo en todo el cual una de las cantidades $\dot{x}$, $\dot{y}$ es diferente de cero. Si, por ejemplo, $\dot{x} \neq 0$, se

puede asignar a α un valor que varíe continuamente con t en el intervalo tomando simplemente

$$\alpha = \alpha(t) = \text{arc tan}\,\frac{\dot{y}}{\dot{x}} + n\pi,$$

donde n es un entero fijo, posiblemente negativo, y "arc tan" indica el valor principal de la función (cfr. p. 236), el cual está entre $-\pi/2$ y $\pi/2$. Análogamente, si $\dot{y} \neq 0$ en el intervalo, se puede tomar para α la expresión

$$\alpha(t) = \text{arc cot}\,\frac{\dot{x}}{\dot{y}} + n\pi = \frac{\pi}{2} - \text{arc tan}\,\frac{\dot{x}}{\dot{y}} + n\pi.$$ [1]

En cualquier caso, se encuentra por derivación directa para cualquier representación paramétrica:

$$\dot{\alpha} = \frac{d\alpha}{dt} = \frac{\dot{x}\ddot{y} - \ddot{x}\dot{y}}{\dot{x}^2 + \dot{y}^2}.$$

Puesto que también (ver (12*a*), p. 371)

$$\dot{s} = \frac{ds}{dt} = \sqrt{\dot{x}^2 + \dot{y}^2},$$

se obtiene para la curvatura $d\alpha/ds = \dot{\alpha}/\dot{s}$ de la curva la expresión

$$\kappa = \frac{d\alpha}{ds} = \frac{\dot{\alpha}}{\dot{s}} = \frac{\dot{x}\ddot{y} - \dot{y}\ddot{x}}{(\dot{x}^2 + \dot{y}^2)^{3/2}}. \tag{15}$$

Escogiendo en particular la longitud de arco s como el parámetro t se tiene

$$\dot{x}^2 + \dot{y}^2 = 1$$

[véase Ec. (13), p. 372] y, por lo tanto, se obtiene el resultado simplificado

$$\kappa = \dot{x}\ddot{y} - \dot{y}\ddot{x}. \tag{15a}$$

Signo y valor absoluto de la curvatura

Introduciendo un nuevo parámetro $\tau = \tau(t)$ en lugar de t no se afecta la dirección de la tangente y, por lo tanto, no ocurre un cambio en α. Análogamente, el valor absoluto de la diferencia de los valores de s en dos puntos tiene un significado geométrico que es independiente de la elección

[1] Podríamos definir $\alpha(t)$ como una función continua para *todos* los valores del parámetro t dividiendo el intervalo paramétrico total en subintervalos en cada uno de los cuales $\dot{x} \neq 0$ o $\dot{y} \neq 0$. En cada subintervalo podemos definir entonces $\alpha(t)$ por medio de una de las expresiones de arriba, eligiendo el entero constante n en cada subintervalo de manera que los valores de α en el extremo común de dos intervalos adyacentes, calculados usando las expresiones correspondientes a esos intervalos, coincidan.

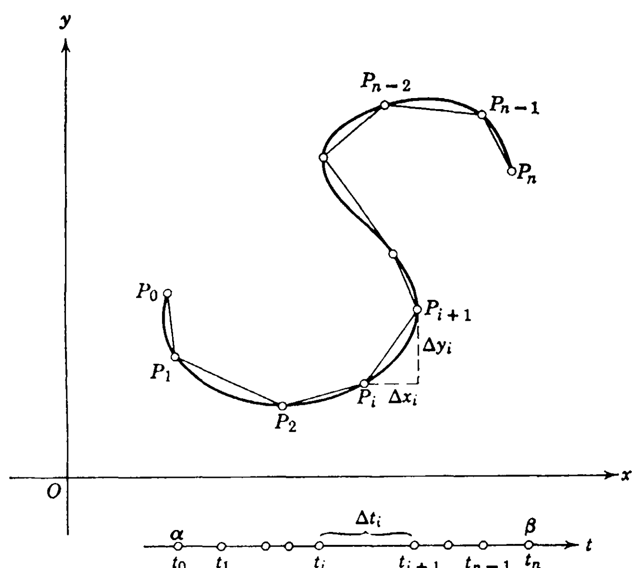


Figura 4.21 Rectificación de curvas.

del parámetro, a saber, el significado de distancia medida a lo largo de la curva. Sin embargo, el *signo* de la diferencia debe siempre tomarse como el signo de la diferencia entre los correspondientes valores paramétricos, puesto que se ha definido s como una función creciente de t. Así, el valor absoluto de la curvatura $|\kappa| = |d\alpha/ds|$ no depende de la elección del pará-

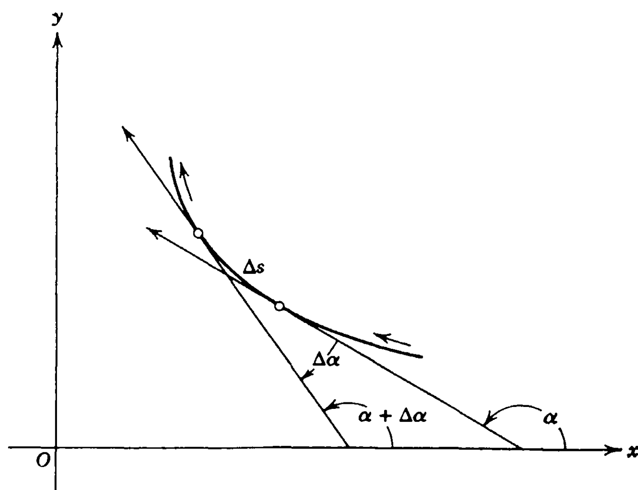


Figura 4.21 (*a*) Curvatura $\kappa = \text{lím}\, \Delta\alpha/\Delta s$ de una curva. (En el caso ilustrado se tiene $\kappa < 0$.)

metro, en tanto que el signo de κ depende del sentido sobre la curva que corresponde a t creciente. Obviamente, $\kappa > 0$ significa que α crece con s, esto es, que la tangente gira contra reloj conforme se recorre la curva con s o t crecientes (ver Fig. 4.21*a*). En este caso la orientación de la curva C es tal que el lado positivo de ella también es su lado "interno", esto es, el lado hacia el cual C se dobla.

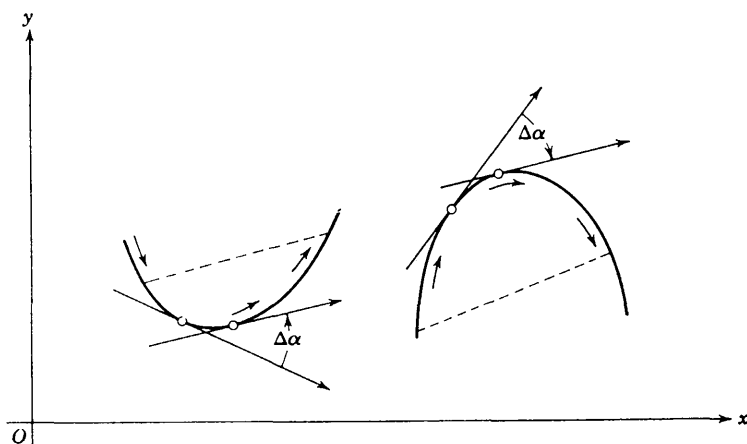


Figura 4.22 Gráfica de una función $f(x)$ convexa (izquierda) y cóncava (derecha).

Si la curva está dada mediante una ecuación $y = f(x)$, se tiene, usando x como parámetro,

$$\kappa = \frac{y''}{(1 + y'^2)^{3/2}}, \tag{16}$$

donde y' e y'' son las derivadas de y con respecto a la variable x. Aquí el signo de la curvatura es el que corresponde a las x crecientes. Obviamente, κ es positiva para $y'' > 0$; en este caso la tangente gira contra reloj conforme x crece; y la función $f(x)$ se llama *convexa.* La porción de la curva que une dos puntos cualesquiera está por debajo de la línea recta que une los mismos puntos. Para $y'' < 0$ la tangente gira en el sentido del reloj para x creciente, y la función f se llama *cóncava.* (Fig. 4.22.) Aquí la curva está por encima de la cuerda que une dos de sus puntos. El caso intermedio en que la curvatura tiene el valor cero corresponde (en términos generales) a un *punto de inflexión,* en el cual $y'' = 0$ (ver p. 258).

Ejemplos. Para la curvatura de la circunferencia de radio a dada por $x = a\cos t$, $y = a \operatorname{sen} t$ se encuentra el valor constante $1/a$ a partir de la fórmula general (15). Así, la curvatura de una circunferencia recorrida en el sentido contra reloj es la recíproca del radio. Este resultado nos asegura que nuestra definición de curvatura es realmente adecuada; pues en

el caso de una circunferencia pensamos como cosa natural que la recíproca del radio sea una medida de su curvatura.

Un segundo ejemplo es la curva definida mediante la función $y = x^3$. La curvatura es

$$\kappa = \frac{6x}{(1 + 9x^4)^{3/2}}.$$

Para $x < 0$ la función $y = x^3$ es cóncava puesto que $\kappa < 0$, y la tangente gira en el sentido del reloj, en tanto que para $x = 0$ se tiene un punto de inflexión, y para $x > 0$ la función se vuelve convexa.

Una función cuya curvatura es idénticamente igual a cero es una línea recta, como puede verse fácilmente por nuestra definición, y la línea recta es la única curva que satisface esa condición.

Circunferencia de curvatura y centro de curvatura

Introducimos $\rho = 1/\kappa$. La cantidad $|\rho| = 1/|\kappa|$ se llama *radio de curvatura* en el punto en cuestión. (Es infinito en un punto de inflexión, en el que $\kappa = 0$.) Para una circunferencia el radio de curvatura en cualquier punto es justamente el radio de la circunferencia.

A cualquier punto $P = (x, y)$ de la curva C se le asigna una circunferencia tangente a C en P y que tiene la misma curvatura que C cuando la curva y la circunferencia se recorren en el mismo sentido en P. Esta circunferencia se llama *circunferencia de curvatura* de la curva C en el punto P.[1] Su centro es el *centro de curvatura* de la curva C correspondiente al punto P. (Fig. 4.23). Puesto que C y el círculo tienen el mismo radio de curvatura, el radio de la circunferencia debe ser el radio de curvatura $|\rho|$ de C, y el centro (ξ, η) de la misma debe estar sobre la normal a C en el punto P y a una distancia $|\rho|$ de este punto. Además, puesto que C y la circunferencia se curvan hacia el mismo lado, el centro está a lo largo de la dirección normal en P, sobre el lado positivo o el negativo según que la curvatura κ sea positiva o negativa.

La dirección que va de P al centro de curvatura forma un ángulo $\alpha + \pi/2$ con el eje x positivo si $\kappa > 0$. Así, si ξ, η son las coordenadas del centro de curvatura y x, y las de P, se tiene [Ver Ec. (7), p. 364]:

$$\frac{\xi - x}{\rho} = \cos\left(\alpha + \frac{\pi}{2}\right) = -\operatorname{sen}\alpha = \frac{-\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}},$$

$$\frac{\eta - y}{\rho} = \operatorname{sen}\left(\alpha + \frac{\pi}{2}\right) = \cos\alpha = \frac{\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2}}.$$

[1] Algunos dicen "*círculo* de curvatura", aunque es claro que se trata de una *circunferencia*. Tal lenguaje erróneo aparece en varias lenguas. *(N. del R.)*

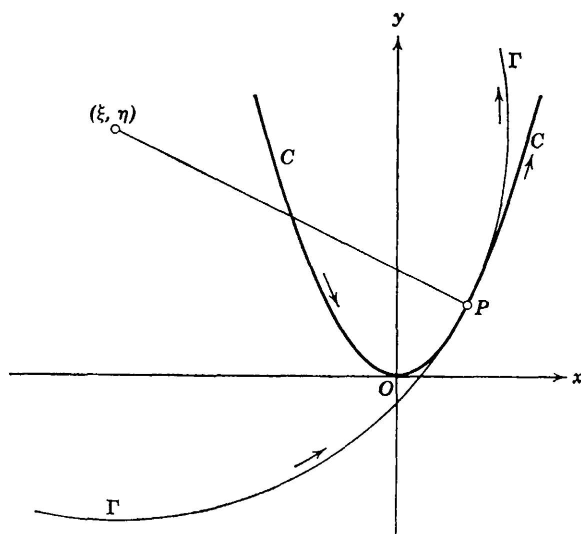


Figura 4.23. Círculo de curvatura Γ y centro de curvatura (ξ, η) correspondientes a un punto P de la curva C.

Por lo tanto, para $\kappa > 0$ resulta

$$(17) \qquad \xi = x - \frac{\rho \dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}}, \qquad \eta = y + \frac{\rho \dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2}}.$$

Si el arco cuya longitud es s se toma como parámetro t, se obtienen las sencillas expresiones

$$(17a) \qquad \xi = x - \rho \dot{y}, \qquad \eta = y + \rho \dot{x}.$$

Las mismas fórmulas para ξ, η se obtienen cuando $\kappa < 0$, en cuyo caso ei radio de curvatura es $-\rho$ y, además, la dirección que va de P al centro forma un ángulo $\alpha - \pi/2$ con el eje x positivo.

La circunferencia de curvatura como circunferencia osculatriz

Las fórmulas (17) dan una expresión para el centro de curvatura en términos del parámetro t correspondiente al punto P de la curva. Conforme t recorre todos los valores del intervalo paramétrico, el centro de curvatura describe una curva, la llamada *evoluta* de la curva dada; y, puesto que, junto con x e y, se tienen que considerar $\dot{x}$, $\dot{y}$ y ρ como funciones conocidas de t, las fórmulas anteriores dan las ecuaciones paramétricas de esta evoluta. En el Apéndice I, p. 442, se encontrarán ejemplos y una discusión de las propiedades geométricas de la evoluta.

Dos curvas cualesquiera se dice que *"osculan"* en un punto P, o que tienen *"contacto de orden dos"* en P, si dichas curvas pasan por P, tienen

la misma tangente en P y tienen también la misma curvatura cuando se orientan de la misma manera. Obviamente, dos curvas osculantes tienen la misma circunferencia y el mismo centro de curvatura en P. Si las curvas están dadas mediante ecuaciones $y = f(x)$ e $y = g(x)$ en forma no paramétrica, es fácil expresar la condición de que ellas tengan un punto de contacto P y las mismas tangentes y curvaturas en ese punto. Si x es la abscisa del punto de contacto P, se tiene $f(x) = g(x)$, $f'(x) = g'(x)$; y la igualdad de las curvaturas es expresada mediante

$$\frac{f''(x)}{[1 + f'^2(x)]^{3/2}} = \frac{g''(x)}{[1 + g'^2(x)]^{3/2}},$$

y, por lo tanto, también es $f''(x) = g''(x)$. Así, la condición para que exista un punto de contacto con curvaturas iguales es que los valores de f y g, junto con los valores de las derivadas primera y segunda, coincidan en ese punto.

Considérense una curva C: $y = f(x)$ y su circunferencia de curvatura Γ en P representada mediante $y = g(x)$ en una vecindad de P. Puesto que la circunferencia Γ coincide con su circunferencia de curvatura, vemos que C y Γ tienen la misma circunferencia de curvatura y, por lo tanto, osculan en P. Consecuentemente, en el punto de contacto es $f(x) = g(x)$, $f'(x) = g'(x)$, $f''(x) = g''(x)$. Se dice que esta circunferencia es "la circunferencia que más se aproxima" a la curva en el punto P de contacto, puesto que ninguna otra circunferencia que intersecta a la curva en el punto de contacto tiene "contacto de orden dos" con C en ese punto. La circunferencia de curvatura es la *circunferencia osculante* u *osculatriz*. (Véase también capítulo 5, p. 478.)

Por otra parte, de la misma manera que la tangente a una curva es el límite, para $P_1 \to P$, de una recta que pasa por dos puntos consecutivos P y P_1 de C, se puede demostrar que la circunferencia de curvatura en P es el límite de las circunferencias que pasan por tres puntos P, P_1, P_2, para $P_1 \to P$ y $P_2 \to P$. La demostración se deja al lector. (Ver problema 4, p. 455.)

i. Cambio de ejes de coordenadas. Invariancia

Las propiedades inherentes a una situación geométrica o física no dependen del sistema de coordenadas específico o "marco de referencia" con respecto al cual dichas propiedades son formuladas. El carácter intrínseco de propiedades tales como distancia, longitud o ángulo debe reflejarse en enunciados que muestren que las respectivas fórmulas permanecen inalterables o son *invariantes* si se pasa de un sistema de coordenadas a otro. Algunos comentarios breves referentes a este punto son oportunos en esta sección.

Se usarán las ecuaciones generales que relacionan las coordenadas x, y de un punto P en un sistema de coordenadas con las coordenadas ξ, η del mismo punto P en cualquier otro sistema. La posición relativa del segundo conjunto de ejes de coordenadas respecto al primer conjunto está caracterizada mediante las coordenadas a, b que el origen del segundo sistema tiene respecto al primer sistema, y mediante el ángulo γ que el eje positivo ξ forma con el eje x positivo.[1] Las coordenadas (x, y) y (ξ, η) del mismo punto en los dos sistemas están relacionadas (cfr. Fig. 4.24) por medio de la transformación

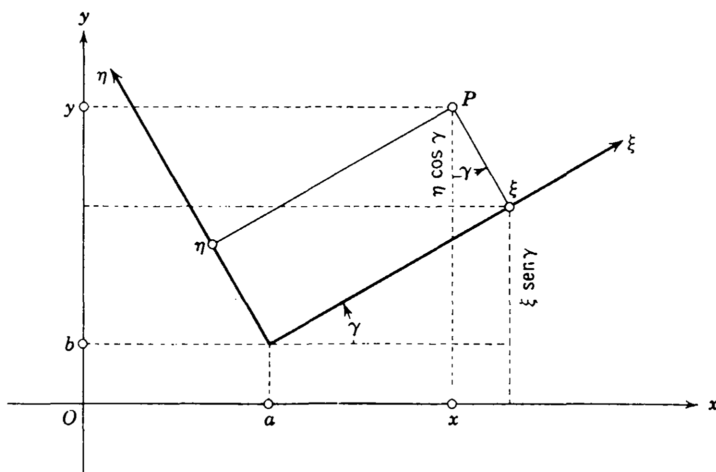


Figura 4.24 Cambio de ejes coordenados.

$$
(18) \qquad \begin{aligned} x &= \xi \cos\gamma - \eta \operatorname{sen}\gamma + a, \\ y &= \xi \operatorname{sen}\gamma + \eta \cos\gamma + b. \end{aligned}
$$

Para $\gamma = 0$ no se tiene rotación de ejes sino solamente un *desplazamiento paralelo* o *traslación,* y las fórmulas toman la forma sencilla $x = \xi + a$, $y = \eta + b$.

Resolviendo las ecuaciones para ξ, η en términos de x, y se encuentra

$$
(18a) \qquad \begin{aligned} \xi &= (x - a)\cos\gamma + (y - b)\operatorname{sen}\gamma, \\ \eta &= -(x - a)\operatorname{sen}\gamma + (y - b)\cos\gamma. \end{aligned}
$$

Si x e y son funciones de un parámetro t que define una curva, a partir de estas fórmulas se obtienen inmediatamente expresiones para ξ y η como funciones de t, dando la representación paramétrica de la misma curva

[1] Nos restringimos a sistemas coordenados "dextrógiros", en los cuales la dirección positiva del segundo sistema de ejes se obtiene mediante una rotación de 90° contra reloj del primer sistema.

en el sistema ξ, η. Derivando con respecto a t (las cantidades a, b, γ, que fijan la posición relativa entre los sistemas de coordenadas, no dependen de t) se obtiene la transformación de las "componentes de velocidad", esto es, la transformación para las derivadas de las coordenadas con respecto a t:

$$\dot{x} = \dot{\xi} \cos \gamma - \dot{\eta} \operatorname{sen} \gamma, \qquad \dot{y} = \dot{\xi} \operatorname{sen} \gamma + \dot{\eta} \cos \gamma.^{1}$$

Se confirma que

$$\dot{x}^2 + \dot{y}^2 = \dot{\xi}^2 + \dot{\eta}^2.$$

Entonces, la expresión $\sqrt{\dot{x}^2 + \dot{y}^2}$ *tiene el mismo valor en todos los sistemas de coordenadas.* Esta propiedad de invariancia es, por supuesto, obvia a partir de la interpretación de esta cantidad como la medida de cambio respecto a t de la longitud a lo largo de la curva, ds/dt. El lector puede verificar mediante un cálculo sencillo que también la expresión $\kappa = (\dot{x}\ddot{y} - x\dot{y})(\dot{x}^2 + y^2)^{-3/2}$ para la *curvatura es invariante.* (Esto, por supuesto, se obtiene también directamente a partir del hecho de que los ángulos que la tangente forma con los ejes ξ y x, respectivamente, difieren sólo en el valor constante γ, de manera que $\kappa = d\alpha/ds$ no puede cambiar.)

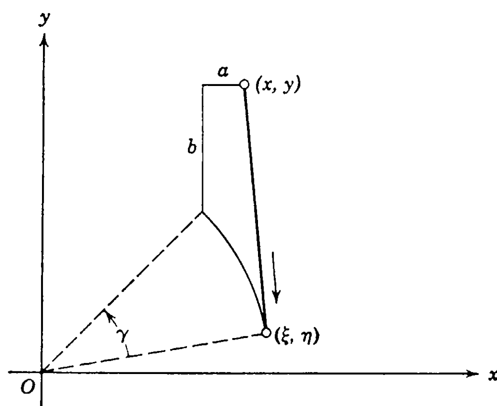


Figura 4.25 Desplazamiento del punto P de la posición (x, y) a la posición (ξ, η).

Las ecuaciones (18) que relacionan las coordenadas x, y con las coordenadas ξ, η son a menudo interpretadas en una forma diferente que corresponde a un desplazamiento. En esta interpretación, los *punto P* son des-

[1] En algunas aplicaciones físicas, donde t representa el tiempo, la posición relativa de los dos sistemas de coordenadas también depende de t. Supóngase que x, y representan las coordenadas de una partícula en un sistema coordenado en reposo, en tanto que ξ, η son las coordenadas de la misma partícula respecto a un sistema móvil, por ejemplo, uno sujeto a la Tierra. Las funciones $x(t)$, $y(t)$ describen la trayectoria de la partícula según la ve un observador en reposo, en tanto que $\xi(t)$, $\eta(t)$ describen la trayectoria vista por un observador móvil. Las fórmulas que relacionan $\dot{x}$, $\dot{y}$ con $\dot{\xi}$, $\dot{\eta}$ tienen entonces que incluir también los términos que provienen de derivar a, b y γ.

plazados en vez de serlo los ejes coordenados (Fig. 4.25). Solamente se usa un sistema de coordenadas. El punto con coordenadas (x, y) en ese sistema es transformado en el punto con coordenadas (ξ, η) del mismo sistema. La *invariancia de la longitud* o de la *curvatura* de una curva significan ahora que esas cantidades no cambian cuando la curva completa sufre un *movimiento rígido.*

*j. Movimiento uniforme en la teoría especial de la relatividad

Como se señaló en la p. 255, existen profundas analogías entre las funciones trigonométricas y las hiperbólicas, que tienen su contraparte geométrica en la correspondencia existente entre las propiedades de las elipses y de las hipérbolas. La relación se hará clara cuando se esté en condiciones de definir las funciones trigonométricas para un argumento imaginario y verificar que $\cos(it) = \cosh t$, $\operatorname{sen}(it) = i \operatorname{senh} t$ (sección 7.7*a*). Como una aplicación de esta analogía, considérense las "rotaciones hiperbólicas" del plano, que pueden ser identificadas con las *transformaciones de Lorentz* de una recta en la teoría especial de la relatividad de Einstein.

Se encontró en (18*a*) de p. 379 que una rotación de los ejes de coordenadas en un ángulo γ, en la que el origen queda fijo, puede describirse mediante las ecuaciones

$$(18b) \qquad \xi = x\cos\gamma + y \operatorname{sen}\gamma, \qquad \eta = -x\operatorname{sen}\gamma + y\cos\gamma,$$

que relacionan las coordenadas x, y de un punto P en el primer sistema con sus coordenadas ξ, η en el segundo sistema. La distancia de P al origen está dada por la misma expresión en ambos sistemas:

$$OP = \sqrt{x^2 + y^2} = \sqrt{\xi^2 + \eta^2}.$$

Esto surge también inmediatamente a partir de las ecuaciones de transformación si se hace uso de la identidad $\cos^2\gamma + \operatorname{sen}^2\gamma = 1$.

Se considera ahora la transformación análoga con coeficientes que son funciones hiperbólicas en lugar de funciones trigonométricas:

$$(19) \qquad \xi = x\cosh\alpha - t\operatorname{senh}\alpha, \qquad \tau = -x\operatorname{senh}\alpha + t\cosh\alpha.$$

Estas fórmulas pueden obtenerse a partir de las fórmulas (18*b*) para rotaciones, tomando para el ángulo de rotación γ y las coordenadas y y η cantidades imaginarias puras:

$$\gamma = i\alpha, \qquad y = it, \qquad \eta = i\tau.$$

Se observa que para un valor real de α (que significaría un ángulo imaginario de rotación γ en la interpretación original) las fórmulas (19) definen ξ y τ como funciones lineales reales de x y t. Estas funciones tienen la propiedad especial de que

$$\begin{aligned}\xi^2 - \tau^2 &= (x\cosh\alpha - t\operatorname{senh}\alpha)^2 - (-x\operatorname{senh}\alpha + t\cosh\alpha)^2\\ &= x^2 - t^2,\end{aligned}$$

como una consecuencia de la identidad $\cosh^2\alpha - \operatorname{senh}^2\alpha = 1$. (Esto surge, por supuesto, también a partir de la observación de que $x^2 - t^2 = x^2 + y^2$ es el cuadrado de la distancia al origen en el plano x, y.) Ahora interprétese t como el tiempo y x como una coordenada espacial que describe la posición de un punto en un espacio unidimensional, esto es, en una línea recta. Cualquier evento ocurre en un cierto punto y en un cierto tiempo. Estas dos partes de la información las proporcionan los dos números x y t que dan respectivamente la distancia x (con su signo) del punto al origen 0 y el tiempo t que ha transcurrido a partir del instante $t = 0$. En la teoría de la relatividad se adopta el punto de vista de que los valores medidos de

esta distancia y los tiempos transcurridos dependen del sistema de referencia utilizado por el observador, esto es, del sistema de coordenadas especial en el continuo espacio-tiempo. Las cantidades ξ, τ obtenidas a partir de las fórmulas (19) describirán el mismo evento en un sistema de referencia diferente en el cual las distancias e intervalos de tiempo pueden tener valores también diferentes. La cantidad que permanece inalterable al pasar de un sistema de referencia a otro (lo que se conoce como "transformación de Lorentz") es

$$\sqrt{x^2 - t^2} = \sqrt{\xi^2 - \tau^2},$$

la "distancia espacio-tiempo" del evento al origen. Para un observador que use el segundo sistema, la cantidad ξ es la distancia espacial medida a partir del origen $\xi = 0$. Ese origen es un punto para el cual

$$x \cosh \alpha - t \operatorname{senh} \alpha = 0,$$

o bien $x/t = \tanh \alpha$. Así, el origen del segundo sistema es un punto que en el primer sistema parece moverse con velocidad uniforme $v = dx/dt = \tanh \alpha$ relativa al origen de este sistema. Por lo tanto, la transformación de Lorentz relaciona los valores de distancias y tiempos según los registran observadores en los dos sistemas que se mueven con velocidad relativa constante v. Aquí

$$v = \frac{\operatorname{senh} \alpha}{\cosh \alpha} = \frac{e^{\alpha} - e^{-\alpha}}{e^{\alpha} + e^{-\alpha}}$$

está comprendida necesariamente entre -1 y $+1$, de manera que las velocidades relativas de los dos sistemas están restringidas a valores numéricamente menores que 1. El valor 1 representa aquí, para una elección adecuada de unidades, la velocidad c de la luz, que no puede ser excedida por v.

Para una u constante la ecuación $x = ut$ corresponde a un punto que en el primer sistema se mueve con velocidad u, partiendo de $x = 0$ en el tiempo $t = 0$. En el segundo sistema el mismo punto tendrá la velocidad

$$\omega = \frac{d\xi}{d\tau} = \left(\frac{d\xi}{dt}\right) \Big/ \left(\frac{d\tau}{dt}\right) = \frac{u - \tanh \alpha}{1 - u \tanh \alpha} = \frac{u - v}{1 - uv}.$$

Este resultado, válido en la teoría especial de la relatividad de Einstein, difiere del que se obtendría en la cinemática clásica, donde la velocidad ω de un punto con respecto a un sistema que se mueve con velocidad v estaría dada simplemente por $\omega = u - v$. La fórmula relativista muestra que $\omega = u$ cuando $u = +1$ o -1; esto corresponde al hecho, sugerido por el famoso experimento de Michelson-Morley, de que la velocidad de la luz es la misma para observadores que se mueven con velocidades diferentes.

k. Integrales que expresan áreas dentro de curvas cerradas

En el capítulo 2 el concepto de integral fue motivado por medio de referencia al "área bajo una curva", esto es, el área de una franja de forma especial. Esta reducción al caso especial de áreas bajo curvas no es completamente satisfactoria, pues las áreas que realmente se encuentran con más frecuencia son las de dominios limitados por curvas cerradas C, los que son de una forma más general que las franjas cuyas áreas pueden representarse mediante integrales de la forma $\int_a^b f(x)\, dx$.

La fórmula fundamental

Se derivará ahora una elegante representación integral de validez general para el área limitada por una curva cerrada C que está dada en forma paramétrica, descomponiendo simplemente el área en cuestión en áreas de franjas especiales. Esta representación será independiente de la representación paramétrica, así como también del sistema de coordenadas. Además, expresará el *área orientada* dentro de la curva de acuerdo con el sentido asignado a la frontera C; esto es, asignará al área dentro de una curva cerrada simple C el signo negativo o el positivo según que la frontera esté orientada en el sentido del reloj o contra reloj.

Supóngase que la curva simple C cerrada y orientada se da mediante $x = x(t)$, $y = y(t)$, donde t varía en el intervalo $\alpha \leq t \leq \beta$ y el sentido de t creciente determina el sentido de C. Supóngase también que x e y son funciones continuas de t (con el mismo valor en $t = \alpha$ que en $t = \beta$) y que sus primeras derivadas $\dot{x}$ e $\dot{y}$ son continuas, con la posible excepción de un número finito de discontinuidades de salto en el caso de que C tenga esquinas. Con estas hipótesis se demostrará la fórmula fundamental

$$(20) \qquad A = -\int_\alpha^\beta y\dot{x}\,dt = \int_\alpha^\beta x\dot{y}\,dt = \frac{1}{2}\int_\alpha^\beta (x\dot{y} - y\dot{x})\,dt$$

para el área orientada A encerrada por C.

Que las tres representaciones integrales en la fórmula son equivalentes se sigue de inmediato integrando por partes la primera de ellas y usando las condiciones de periodicidad $x(\alpha) = x(\beta)$, $y(\alpha) = y(\beta)$. La tercera representación, más simétrica, es justamente la media aritmética de las dos primeras.

La expresión (20) no depende de la posición del sistema de coordenadas en el plano. En efecto, la expresión simétrica

$$A = \frac{1}{2}\int_\alpha^\beta (x\dot{y} - y\dot{x})\,dt$$

muestra claramente que el valor de A es independiente de la elección del sistema de coordenadas. Como se vio en la p. 379, un cambio de coordenadas de un sistema xy a un sistema $\xi\eta$ se realiza mediante una substitución de la forma

$$x = \xi\cos\gamma - \eta\,\mathrm{sen}\,\gamma + a,$$
$$y = \xi\,\mathrm{sen}\,\gamma + \eta\cos\gamma + b,$$

con a, b y γ constantes. Derivando estas fórmulas respecto a t se llega a

$$\dot{x} = \dot{\xi}\cos\gamma - \dot{\eta}\,\mathrm{sen}\,\gamma, \qquad \dot{y} = \dot{\xi}\,\mathrm{sen}\,\gamma + \dot{\eta}\cos\gamma$$

y, en consecuencia,

$$x\dot{y} - y\dot{x} = \xi\dot{\eta} - \eta\dot{\xi} + a\dot{y} - b\dot{x}.$$

Por tanto, la expresión $x\dot{y} - y\dot{x}$ es invariante bajo rotaciones alrededor del origen (esto es, cuando $a = b = 0$). Aún cuando a y b no se anulen, el valor de la integral para A no resulta afectado ya que

$$\int_{\alpha}^{\beta} (a\dot{y} - b\dot{x})\, dt = (ay - bx)\Big|_{\alpha}^{\beta} = 0$$

para la curva cerrada C.

Prueba de la fórmula fundamental (20). *Integrales de línea sobre arcos simples.* La fórmula básica (20) se prueba con algunos pasos sencillos.

Primero, sea C un arco simple *orientado,* con punto inicial P_0 y punto final P_1. Sea $x = x(t)$, $y = y(t)$ cualquier representación paramétrica de C con P_0 y P_1 correspondiendo respectivamente a $t = t_0$, y $t = t_1$. (Aquí t_0 puede ser mayor o menor que t_1.) Entonces la integral

$$A = -\int_{t_0}^{t_1} y \frac{dx}{dt}\, dt$$

depende solamente de C y no de la representación paramétrica particular. Esta es una consecuencia obvia de la regla de substitución. Si se introduce un nuevo parámetro τ mediante la función monótona $\tau = \chi(t)$, donde $\tau_0 = \chi(t_0)$, $\tau_1 = \chi(t_1)$, la integral correspondiente es[1]

$$-\int_{\tau_0}^{\tau_1} y \frac{dx}{d\tau}\, d\tau = -\int_{t_0}^{t_1} y \frac{dx}{d\tau} \frac{d\tau}{dt}\, dt = -\int_{t_0}^{t_1} y \frac{dx}{dt}\, dt = A.$$

Queda, por consiguiente, justificado el omitir en la expresión para la integral A la referencia a cualquier parámetro especial t, pudiendo escribirse simplemente

$$A = A_C = -\int_C y\, dx.$$

Aquí, A_C para un arco simple y orientado C debe calcularse refiriendo el arco a un parámetro t, usando $dx = (dx/dt)\, dt$ y tomando como límites para la integración sobre t los valores paramétricos correspondientes a los puntos extremos de C en el orden determinado por la orientación del arco.*

Si C' es el arco que se obtiene de C cambiando la orientación de éste, esto es, el arco con punto inicial P_1 y punto final P_0, usando la misma representación paramétrica para C' se tiene

[1] Suponemos continuas no sólo $x(t)$ e $y(t)$ sino también $\tau(t)$ y sus derivadas, con la posible excepción, para estas últimas, de un número finito de discontinuidades de salto.

* La integral $\int_C y\, dx$ es un ejemplo de las *integrales de línea* generales $\int_C p\, dx + q\, dy$, que serán discutidas en el Volumen II.

$$A_{C'} = -\int_{t_1}^{t_0} y\frac{dx}{dt}dt = +\int_{t_0}^{t_1} y\frac{dx}{dt}dt = -A_C.$$

Por lo tanto, cambiando la orientación del arco se cambia el signo de la integral A_C.

Si el arco simple y orientado C se divide en subarcos orientados C_1, C_2, ..., C_n, cada uno con la misma orientación que C, se tendrá obviamente

$$A_C = A_{C_1} + A_{C_2} + \cdots + A_{C_n}.$$

Pues, en efecto, en una representación paramétrica de C en la que, por ejemplo, el sentido de C sea el de t creciente, esta descomposición corresponde a una subdivisión del intervalo paramétrico $t_0 \leq t \leq t_n$ de C en subintervalos $t_0 \leq t \leq t_1$, $t_1 \leq t \leq t_2$, ..., $t_{n-1} \leq t \leq t_n$ correspondientes a C_1, ..., C_n. El resultado se obtiene entonces por la aditividad de las integrales.

La aditividad de las integrales A_C facilita mucho el cálculo de A_C en casos en que C consiste de varios arcos suaves C_1, C_2..., cada uno con su propia representación paramétrica. No se necesita construir artificialmente una representación paramétrica única para la curva C completa, sino calcular cada A_{C_i} separadamente a partir de su representación paramétrica y luego efectuar la suma. Además, los A_{C_i} pueden sumarse en cualquier orden; sólo se necesita asegurarse de que todos los C_i tienen la misma orientación que C.

La integral de línea fundamental para curvas cerradas

Se puede ahora definir A_C para cualquier curva simple C, cerrada y orientada, descomponiendo C en arcos simples C_1, ..., C_n con orientaciones que coincidan con la de C y formando la suma de los A_{C_i}.[1] Si la curva C completa tiene la representación paramétrica $x = x(t)$, $y = y(t)$, para $\alpha \leq t \leq \beta$, donde el sentido de las t crecientes da la orientación de C y donde $t = \alpha$ y $t = \beta$ corresponden al mismo punto, entonces A_C está dada nuevamente por

$$A_C = -\int_{\alpha}^{\beta} y\frac{dx}{dt}dt.$$

De la misma manera se puede definir A_C para curvas orientadas C no simples mediante la descomposición en arcos simples orientados, aún cuando C consista de varias partes ajenas pero con la condición de que cada porción de C tenga un sentido definido.

[1] Que el valor de A_C así obtenido no depende de la manera particular en que C es dividida en arcos simples, se demuestra fácilmente: primero, la propiedad aditiva de A para arcos simples muestra que el refinamiento de una subdivisión dada por la adición de puntos de división no cambia el valor resultante de A_0; además, dos subdivisiones cualesquiera pueden reemplazarse por aquélla que es un refinamiento de ambas, sin que cambie el valor de A_C.

La integral fundamental como un área

Regresamos ahora al punto principal; esto es, identificamos las expresiones A_C para una curva cerrada con la intuitiva cantidad geométrica que representa el área orientada que C encierra.

Consideramos primero un dominio G limitado en su parte superior por un arco C_1: $y = g(x)$ para $a \leq x \leq b$; en su parte inferior por un arco C_3: $y = f(x)$ para $a \leq x \leq b$; y a los lados por los segmentos de recta C_2,

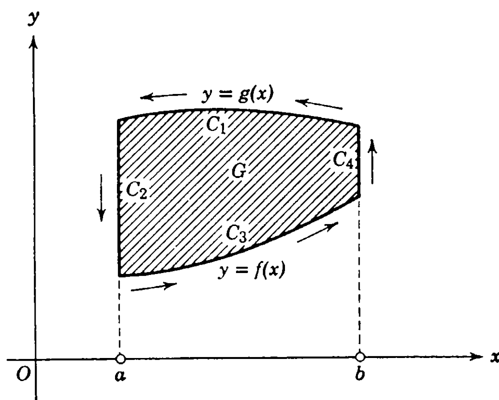


Figura 4.25(a) Area de una "celda".

C_4 dados por $x = a$ y $x = b$ (Fig. 4.25a). Aquí se permite que C_2 y C_4 se reduzcan a puntos. Si damos a C la orientación contra reloj, el arco C_1 será recorrido en el sentido de las x decrecientes y el arco C_3 en el de las x crecientes. Al formar A_C como la suma de las cuatro A_{C_i}, las porciones C_2 y C_4 a lo largo de las cuales x es constante no aportan contribución alguna pues $dx/dt = 0$. Usando x como parámetro sobre los arcos C_1 y C_3 encontramos

$$A_C = A_{C_1} + A_{C_3} = -\int_b^a g(x)\,dx - \int_a^b f(x)\,dx$$
$$= \int_a^b g(x)\,dx - \int_a^b f(x)\,dx.$$

Esto es, claramente, el área positiva del dominio G en el caso en que G quede completamente por encima del eje x, siendo igual a la diferencia entre las áreas que están respectivamente por debajo de las curvas C_1 y C_3. Podemos siempre garantizar que G queda por encima del eje x reemplazando simplemente y por $y + c$, donde c es una constante adecuada, esto es, mediante una traslación en la dirección y. Esto no altera las áreas y tampoco afecta el valor $A_C = -\int_C y\,dx$ para una curva cerrada C,

como vimos antes. Por lo tanto, para dominios G del tipo descripto que tengan fronteras C intersectadas a lo más en dos puntos por paralelas al eje y, la integral A_C representa el área, tomada positiva si C está orientada contra reloj y negativa para la orientación opuesta. El mismo resultado se obtiene para áreas limitadas por curvas C que sean intersectadas a lo más en dos puntos por paralelas al eje x. Lo único que se necesita es escribir A_C en la forma $\int_C x\,dy$ e intercambiar x e y en el argumento anterior. A los dominios G pertenecientes a uno de estos dos tipos los llamaremos "celdas". Hablaremos de "celdas orientadas" cuando las curvas de frontera correspondientes tengan una orientación o la otra.

Consideremos ahora un dominio G con la frontera orientada C y que esté compuesto por cierto número de celdas simples $G_1, G_2, \ldots, G_n$ con fronteras $C_1, \ldots, C_n$ respectivamente; y se supone que todas estas celdas tienen la misma orientación, digamos contra reloj. Entonces, como se indica en la Fig. 4.26, las partes de las fronteras que son comunes a dos celdas adyacentes se recorren en sentidos diferentes al considerar sucesivamente dichas partes como arcos de frontera de una celda y de la celda vecina Por consiguiente, si sumamos las integrales $A_{C_i} = -\int y\,dx$ que se tienen para las diferentes celdas las contribuciones de todas las fronteras interiores de las celdas se anulan, y así obtenemos

$$A = \sum_{i=1}^{n} A_{C_i} = \sum_{i=1}^{n} \left(-\int_{C_i} y\,dx \right) = -\int_C y\,dx = A_C,$$

donde A es el área orientada del dominio total G.

Así, las fórmulas (20) para el área A de un dominio orientado G interior a una curva cerrada quedan demostradas para todos los dominios que puedan descomponerse en celdas simples, por ejemplo, trazando paralelas a los ejes coordenados.

Para todos los dominios que encontraremos, esta hipótesis se satisfará obviamente; como ocurre, por ejemplo, para dominios poligonales.

Comentarios adicionales

Finalmente, podría agregarse que de la misma manera se demuestra la validez de la fórmula para el área aún en el caso de dominios múltiplemente conexos, tales como los dominios en forma de anillo, que puedan descomponerse en un número finito de celdas simples. En estos casos todas las curvas de frontera tienen que ser recorridas sistemáticamente en un sentido tal que el interior de G esté siempre en el lado "izquierdo" o siempre en el lado "derecho".

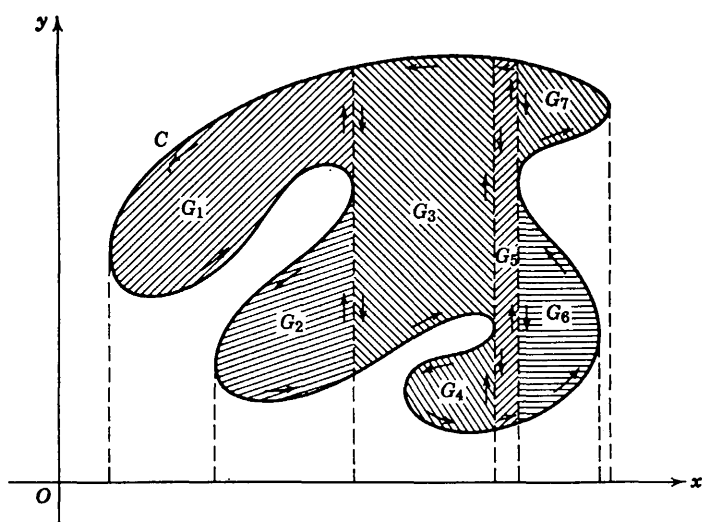


Figura 4.26 Descomposición de un dominio orientado en celdas orientadas.

Las fórmulas para A conservan su validez aún cuando C no sea una curva simple sino una que se intersecta a sí misma dividiendo el plano en más de dos regiones. En este caso podemos considerar la fórmula como una guía para interpretar adecuadamente el área como una combinación aditiva de las áreas orientadas que corresponden a las diferentes partes conexas del plano limitadas por C. Este asunto se discutirá en el Apéndice II de este capítulo.

Ejemplos. Como un ejemplo, podemos encontrar el *área encerrada por la elipse* $x^2/a^2 + y^2/b^2 = 1$. Usando la orientación contra reloj para la elipse y a partir de la representación paramétrica

$$x = a\cos t, \qquad y = b \operatorname{sen} t, \qquad \text{para} \quad 0 \leq t \leq 2\pi,$$

encontramos que

$$A = \frac{1}{2}\int_0^{2\pi} (x\dot{y} - y\dot{x})\, dt = \frac{1}{2}\int_0^{2\pi} ab\, dt = \pi ab.$$

Area en coordenadas polares. Para expresar el área en coordenadas polares r y θ consideremos primero el área A de la región limitada por un segmento de curva $r = f(\theta)$ y los radios $\theta = \alpha$ y $\theta = \beta$. Suponemos que $\alpha < \beta$ y que θ puede usarse como parámetro a lo largo de la curva (esto es, que puntos diferentes tienen ángulos polares diferentes). Usamos para A la expresión

$$A = \frac{1}{2}\int (x\, dy - y\, dx) = \frac{1}{2}\int (x\dot{y} - y\dot{x})\, dt.$$

la cual tiene entonces que ser extendida sobre la parte curva de la frontera y sobre los dos radios. Sobre los radios $\theta = \alpha$ y $\theta = \beta$ podemos usar r como parámetro y, a partir de $x = r\cos\theta$, $y = r\operatorname{sen}\theta$ y $\theta =$ constante, encontrar que $\dot{x} = \cos\theta$, $\dot{y} = \operatorname{sen}\theta$, de manera que $x\dot{y} - y\dot{x} = 0$. Sobre la parte curva usamos θ como parámetro. Entonces,

$$\dot{x} = \frac{dr}{d\theta}\cos\theta - r\operatorname{sen}\theta, \qquad \dot{y} = \frac{dr}{d\theta}\operatorname{sen}\theta + r\cos\theta,$$

y así $x\dot{y} - y\dot{x} = r^2$. En consecuencia,

$$A = \frac{1}{2}\int_\alpha^\beta r^2\,d\theta = \frac{1}{2}\int_\alpha^\beta f^2(\theta)\,d\theta. \tag{21}$$

Para una curva cerrada simple C que contenga el origen en su interior y sea intersectada por cada rayo que parta del origen en exactamente un punto, podemos usar θ como parámetro, con $0 \leq \theta \leq 2\pi$, y encontrar así para el área encerrada

$$A = \frac{1}{2}\int_0^{2\pi} r^2\,d\theta. \tag{22}$$

La fórmula (21) para el área en coordenadas polares puede también derivarse directamente a partir de la definición de integral. Para ello se divide el dominio en sectores trazando radios desde el origen (Fig. 4.27). Cada sector queda descripto por las desigualdades

$$\theta_{i-1} < \theta < \theta_i, \qquad 0 < r < f(\theta).$$

Obviamente, el área del sector queda entre las áreas de los sectores circulares inscripto y circunscripto; el área de un sector del dominio es entonces igual a $\frac{1}{2}r^2(\theta_i - \theta_{i-1})$, donde r está comprendido entre el valor más grande y el más pequeño de $f(\theta)$ para el intervalo $\theta_{i-1} < \theta < \theta_i$. Conforme se refina la subdivisión la suma de las áreas de los sectores del dominio evidentemente converge a la integral $\dfrac{1}{2}\displaystyle\int_\alpha^\beta r^2\,d\theta$.

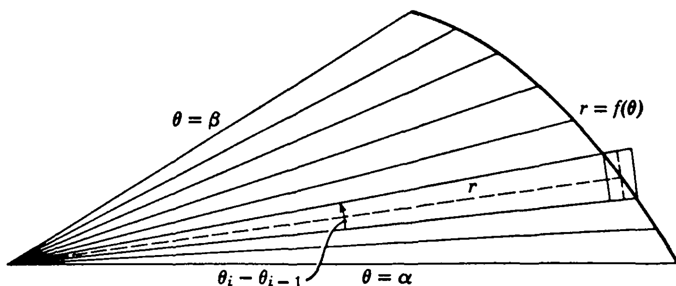


Figura 4.27 Area en coordenadas polares.

Area en un lemniscato

Como un ejemplo de la Ec. (21), consideremos el área limitada por una espira de una leminscata. La ecuación de la lemniscata (cfr. pp. 125-126) es $r^2 = 2a^2 \cos 2\theta$; y una espira es obtenida haciendo variar θ de $-\pi/4$ a $+\pi/4$. Esto da la expresión

$$a^2 \int_{-\pi/4}^{\pi/4} \cos 2\theta \, d\theta = a^2$$

para el área. Por supuesto, la otra espira tiene un área del mismo valor absoluto pero de signo negativo.

Area limitada por un hipérbola

Consideremos ahora el área de un sector limitado por la hipérbola $x^2 - y^2 = 1$, la cual hemos calculado ya en la p. 255 de una manera más bien complicada (véase Fig. 3.12). Para la hipérbola (más bien, para su rama derecha) se tiene la representación paramétrica $x = \cosh t$, $y = \operatorname{senh} t$. Para el doble del área limitada por la hipérbola y los radios que van a los puntos con parámetros 0 y t se encuentra el valor

$$\begin{aligned} 2A &= \int_0^t (x\dot{y} - y\dot{x})\, d\tau = \int_0^t (\cosh^2 \tau - \operatorname{senh}^2 \tau)\, d\tau \\ &= \int_0^t d\tau = t. \end{aligned}$$

(Nuevamente, no hay contribución a la integral por parte de los radios.)

l. Centro de masa y momento de una curva

Veamos ahora algunas ideas que surgen en la mecánica. Consideramos un sistema de n partículas en un plano cuyas masas son $m_1, m_2, \ldots, m_n$ y las respectivas ordenadas $y_1, y_2, \ldots, y_n$. Entonces llamamos a

$$T = \sum_{\nu=1}^{n} m_\nu y_\nu = m_1 y_1 + m_2 y_2 + \cdots + m_n y_n$$

el *momento del sistema de partículas con respecto al eje x*. La expresión $\eta = T/M$, donde M denota la masa total $m_1 + m_2 + \cdots + m_n$ del sistema, define la *altura del centro de masa* del sistema de partículas sobre el eje x, o sea su ordenada. Es, justamente, la *media pesada* de $y_1, y_2, \ldots, y_n$ usando los "factores peso" $m_1, m_2, \ldots, m_n$ (ver. p. 164). Por lo tanto, η es la *altura promedio* de las masas. Análogamente definimos el momento con respecto al eje y y la abscisa del centro de masa.

Podemos extender ahora fácilmente estas definiciones de momento al caso de una curva a lo largo de la cual hay una masa distribuida unifor-

memente, y definir así las coordenadas ξ y η del centro de masa de tal curva. (La hipótesis de una densidad constante, digamos μ, a lo largo de la curva no es esencial: cualquier distribución continua podría discutirse igualmente.)

En un procedimiento típico en la mecánica se comienza con un sistema de un número finito n de partículas y se pasa luego a un límite para $n \to \infty$. Para hacer esto, introducimos la longitud de arco s como un parámetro sobre la curva y subdividimos ésta en arcos de longitudes $\Delta s_1, \Delta s_2, \ldots, \Delta s_n$, usando $(n - 1)$ puntos de división. Representamos la masa $\mu \Delta s_i$ de cada arco Δs_i como si estuviese concentrada en un punto arbitrario del arco, por ejemplo, el punto con ordenada y_i.

Por definición, el momento con respecto al eje x de este sistema de partículas es

$$T = \mu \sum y_i \Delta s_i.$$

Si ahora la mayor de las cantidades Δs_i tiende a cero, esta suma tiende a un límite dado por la integral

$$T = \mu \int_{s_0}^{s_1} y\, ds = \mu \int_{x_0}^{x_1} y\sqrt{1 + y'^2}\, dx, \tag{23}$$

la cual es, por consiguiente, aceptada en forma natural como la definición del momento de la curva con respecto al eje x. Puesto que la masa total de la curva es igual a su longitud multiplicada por μ, es decir,

$$\mu \int_{s_0}^{s_1} ds = \mu(s_1 - s_0),$$

llegamos inmediatamente a las siguientes expresiones para las coordenadas del centro de masa de la curva:

$$\xi = \frac{\int_{s_0}^{s_1} x ds}{s_1 - s_0}, \quad \eta = \frac{\int_{s_0}^{s_1} y\, ds}{s_1 - s_0}. \tag{24}$$

Estos asertos son realmente *definiciones* del momento y del centro de masa de una curva; pero son extensiones tan directas del caso más simple de un número finito de partículas, que uno naturalmente espera —como es, precisamente, el caso— que cualquier afirmación en mecánica sobre el centro de masa o el momento de un sistema de partículas sea válida también para distribuciones continuas de masa a lo largo de curvas.

m. Area y volumen de una superficie de revolución

Regla de Guldin

Si se hace girar una curva $y = f(x)$, para la cual $f(x) \geq 0$, alrededor del eje x, se describe una superficie llamada *superficie de revolución*. El

área de esta superficie, cuyas abscisas suponemos comprendidas entre las cotas x_0 y $x_1 > x_0$, se obtiene mediante una discusión análoga a la anterior. En efecto, si reemplazamos la curva por un polígono inscripto, en vez de una superficie curva tendremos una figura compuesta por un cierto número de conos truncados. La intuición sugiere que podríamos definir el área de la superficie de revolución como el límite de las áreas de estas superficies cónicas cuando la longitud del lado mayor del polígono inscripto tiende a cero. De la geometría elemental sabemos que el área de cada cono truncado es igual a la longitud de la generatriz multiplicada por el perímetro de la sección circular de radio medio. (Fig. 4.28.) Si sumamos estas expresiones y después realizamos el paso al límite obtenemos para el área la expresión

$$(25)\quad A = 2\pi \int_{s_0}^{s_1} y\,ds = 2\pi \int_{x_0}^{x_1} y\sqrt{1+y'^2}\,dx = 2\pi\eta(s_1 - s_0).$$

Expresado en palabras, este resultado dice que el área de una superficie de revolución es igual a la longitud de la curva generadora multiplicada por la distancia recorrida por el centro de masa *(regla de Guldin).*

Del mismo modo encontramos que el volumen interior a la superficie de revolución, limitado en los extremos por los planos $x = x_0$ y $x = x_1 > x_0$, está dado por la expresión

$$(26)\qquad V = \pi \int_{x_0}^{x_1} y^2\,dx.$$

Esta fórmula se obtiene siguiendo la idea intuitiva de que el volumen en cuestión es el límite de los volúmenes de las figuras antes mencionadas consistentes en conos truncados. El resto de la demostración se deja al lector.

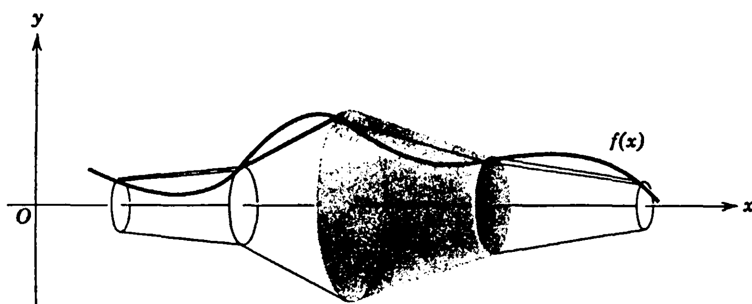


Figura 4.28 Area de una superficie de revolución.

n. Momento de inercia

En el estudio de la rotación de un objeto juegan un papel importante ciertas cantidades llamadas *momentos de inercia.* Estas serán mencionadas aquí brevemente.

Supongamos que una partícula de masa m a una distancia y del eje x gira uniformemente alrededor de ese eje con velocidad angular ω (esto es, en la unidad de tiempo gira un ángulo ω). La *energía cinética* de la partícula, expresada como la mitad del producto de la masa por el cuadrado de la velocidad, es obviamente

$$\frac{m}{2}(y\omega)^2:$$

Al coeficiente de $\frac{1}{2}\omega^2$, esto es, a la cantidad my^2 le llamamos el *momento de inercia de la partícula alrededor del eje x.*

Análogamente, si tenemos n partículas con masas $m_1, m_2, \ldots, m_n$ y ordenadas $y_1, y_2, \ldots, y_n$, llamamos a la expresión

$$T = \sum_i m_i y_i^2$$

el *momento de inercia del sistema de masas* alrededor del eje x. El momento de inercia es una cantidad que pertenece al propio sistema de masas, sin referencia a su estado de movimiento. Su importancia radica en el hecho de que en la rotación rígida del sistema alrededor de un eje, sin que cambie la distancia mutua entre pares de partículas, la energía cinética se obtiene multiplicando el momento de inercia alrededor de ese eje por la mitad del cuadrado de la velocidad angular. Así, el momento de inercia alrededor de un eje juega el mismo papel en la rotación alrededor de ese eje que la masa en el movimiento rectilíneo.

Supóngase que se tiene ahora una curva arbitraria $y = f(x)$ comprendida entre las abscisas x_0 y $x_1 > x_0$, a lo largo de la cual una masa está distribuida uniformemente con densidad unitaria. Para definir el momento de inercia de esta curva se procede justamente como en la sección anterior, llegándose a la siguiente expresión para el momento de inercia alrededor del eje x

$$(27) \qquad T_x = \int_{s_2}^{s_1} y^2\,ds = \int_{s_0}^{x_1} y^2\sqrt{1 + y'^2}\,dx.$$

Para el momento de inercia alrededor del eje y se tiene, en forma semejante,

$$(28) \qquad T_y = \int_{s_0}^{s_1} x^2\,ds = \int_{x_0}^{x_1} x^2\sqrt{1 + y'^2}\,dx.$$

4.2 Ejemplos

De la gran variedad de curvas planas escogemos unos pocos ejemplos típicos para ilustrar los conceptos discutidos.

a. La cicloide común

De las ecuaciones $x = a(t - \operatorname{sen} t)$, $y = a(1 - \cos t)$ (cfr. (1), p. 347) se obtiene $\dot{x} = a(1 - \cos t)$, $\dot{y} = a \operatorname{sen} t$, y para la longitud de arcos se tiene

$$s = \int_0^{\alpha} \sqrt{\dot{x}^2 + \dot{y}^2}\, dt = \int_0^{\alpha} \sqrt{2a^2(1 - \cos t)}\, dt.$$

Puesto que $1 - \cos t = 2 \operatorname{sen}^2 t/2$, el integrando es igual a $2a \operatorname{sen} t/2$ y, por consiguiente, con $0 \leq \alpha \leq 2\pi$, se obtiene

$$s = 2a \int_0^{\alpha} \operatorname{sen}(t/2)\, dt = -4a \cos \frac{t}{2} \Big|_0^{\alpha} = 4a\left(1 - \cos \frac{\alpha}{2}\right) = 8a \operatorname{sen} \frac{\alpha}{4}.$$

Si, en particular, se considera la longitud de arco entre dos cúspides sucesivas, debe tomarse $\alpha = 2\pi$ ya que el intervalo $0 \leq t \leq 2\pi$ de los valores del parámetro corresponde a una revolución del círculo giratorio. Así se obtiene el valor $8a$; esto es, la longitud de arco de la cicloide entre dos cúspides sucesivas es igual a *cuatro* veces el diámetro del círculo giratorio.

De manera análoga se calcula el área limitada por un arco de la cicloide y el eje x:

$$\begin{aligned} I &= \int_0^{2\pi} y\dot{x}\, dt = a^2 \int_0^{2\pi} (1 - \cos t)^2\, dt \\ &= a^2 \int_0^{2\pi} (1 - 2\cos t + \cos^2 t)\, dt \\ &= a^2 \left(t - 2 \operatorname{sen} t + \frac{t}{2} + \frac{\operatorname{sen} 2t}{4} \right) \Big|_0^{2\pi} = 3a^2\pi. \end{aligned}$$

Esta área es, por consiguiente, tres veces el área del círculo giratorio.

Para el radio de curvatura $|\rho| = 1/|\kappa|$ se tiene, por la Ec. (15) de p. 373,

$$\rho = \frac{(\dot{x}^2 + \dot{y}^2)^{3/2}}{\ddot{y}\dot{x} - \dot{y}\ddot{x}} = -2a\sqrt{2(1 - \cos t)} = -4a \left| \operatorname{sen} \frac{t}{2} \right|.$$

En los puntos $t = 0$, $t = \pm 2\pi, \ldots$ esta expresión tiene el valor cero. Estos puntos son las cúspides; en ellos la cicloide intersecta el eje x en ángulo recto.

El área de la superficie de revolución que se forma al hacer girar un arco de la cicloide alrededor del eje x está dado, según nuestra fórmula (25), p. 392, por

$$\begin{aligned} A &= 2\pi \int_0^{8a} y\, ds = 2\pi \int_0^{2\pi} a(1 - \cos t) \cdot 2a \operatorname{sen} \frac{t}{2}\, dt \\ &= 8a^2\pi \int_0^{2\pi} \operatorname{sen}^3 \frac{t}{2}\, dt = 16a^2\pi \int_0^{\pi} \operatorname{sen}^3 u\, du \end{aligned}$$

$$= 16a^2\pi \int_0^{\pi} (1 - \cos^2 u) \operatorname{sen} u \, du.$$

La última integral puede evaluarse mediante la substitución $\cos u = v$; y se encuentra que

$$A = 16a^2\pi \left(-\cos u + \frac{1}{3}\cos^2 u \right) \Big|_0^{\pi} = \frac{64a^2\pi}{3}.$$

Como ejercicio, el lector puede calcular la altura por encima del eje x, η, del centro de masa de la cicloide, y también el momento de inercia T_x. Los resultados son

$$\eta = \frac{4}{3}a = \frac{A}{2\pi s} \quad \text{y} \quad T_x = \frac{256}{15}a^3.$$

b. La catenaria

La catenaria [1] es la curva definida por la ecuación $y = \cosh x$. La longitud de la catenaria entre las abscisas $x = a$ y $x = b$ es

$$s = \int_a^b \sqrt{1 + \operatorname{sen}^2 x}\, dx = \int_a^b \cosh x \, dx = \operatorname{senh} b - \operatorname{senh} a.$$

Para el área de la superficie de revolución obtenida haciendo girar la catenaria alrededor del eje x, la llamada superficie *catenoide*, se encuentra

$$A = 2\pi \int_a^b \cosh^2 x \, dx = 2\pi \int_a^b \frac{1 + \cosh 2x}{a} dx$$
$$= \pi(b - a + \tfrac{1}{2} \operatorname{senh} 2b - \tfrac{1}{2} \operatorname{senh} 2a).$$

A partir de aquí se obtiene a continuación la altura del centro de masa del arco de a a b:

$$\eta = \frac{A}{2\pi s} = \frac{b - a + \frac{1}{2}\operatorname{senh} 2b - \frac{1}{2}\operatorname{senh} 2a}{2(\operatorname{senh} b - \operatorname{senh} a)}.$$

Finalmente, para la curvatura se tiene

$$\kappa = \frac{y''}{(1 + y'^2)^{3/2}} = \frac{\cosh x}{\cosh^3 x} = \frac{1}{\cosh^2 x}.$$

[1] El nombre proviene del hecho de que una cadena suspendida de sus extremos adopta la forma de esta curva. Es curioso que la misma curva surja en una aplicación física completamente diferente. Una película de jabón, limitada por dos círculos en el espacio que están en planos paralelos y con sus centros sobre la misma perpendicular a los planos, tiene la forma de una superficie de revolución obtenida al hacer girar la catenaria alrededor del eje x.

c. La elipse y la lemniscata

Las longitudes de arco de estas dos curvas no pueden reducirse a funciones elementales, sino que pertenecen a la clase de "integrales elípticas" mencionadas en la p. 318.

Para la elipse $y = (b/a)\sqrt{a^2 - x^2}$ se obtiene

$$s = \frac{1}{a}\sqrt{\frac{a^4 - (a^2 - b^2)x^2}{a^2 - x^2}}\,dx = a\int\sqrt{\frac{1 - \eta^2\xi^2}{1 - \xi^2}}\,d\xi,$$

donde se ha hecho $x/a = \xi$, $1 - b^2/a^2 = \eta^2$. Mediante la substitución $\xi = \operatorname{sen}\phi$ esta integral puede escribirse en la forma

$$s = a\int\sqrt{1 - \eta^2\operatorname{sen}^2\phi}\,d\phi.$$

Aquí, para obtener el semiperímetro de la elipse debe hacerse que x recorra el intervalo de $-a$ a $+a$, que corresponde al intervalo

$$-1 \leq \xi \leq +1 \qquad \text{o} \qquad -\frac{\pi}{2} \leq \phi \leq +\frac{\pi}{2}.$$

Para la lemniscata, cuya ecuación en coordenadas polares r, t es $r^2 = 2a^2\cos 2t$, de manera análoga se obtiene

$$s = \int\sqrt{r^2 + \dot{r}^2}\,dt = \int\sqrt{2a^2\cos 2t + 2a^2\frac{\operatorname{sen}^2 2t}{\cos 2t}}\,dt$$

$$= a\sqrt{2}\int\frac{dt}{\sqrt{\cos 2t}} = a\sqrt{2}\int\frac{dt}{\sqrt{1 - 2\operatorname{sen}^2 t}}$$

Si se introduce $u = \tan t$ como variable independiente en la última integral, se tiene

$$\operatorname{sen}^2 t = \frac{u^2}{1 + u^2}, \qquad dt = \frac{du}{1 + u^2},$$

y, en consecuencia,

$$s = a\sqrt{2}\int\frac{du}{\sqrt{1 - u^4}}.$$

En una espira completa de la lemniscata, u corre de -1 a $+1$, y la longitud de arco es entonces igual a

$$a\sqrt{2}\int_{-1}^{+1}\frac{du}{\sqrt{1 - u^4}}$$

una integral elíptica especial que jugó un papel importante en las investigaciones de Gauss.

4.3 Vectores en dos dimensiones

Para las discusiones de curvas planas y muchos otros tópicos en geometría, mecánica y física, la notación vectorial constituye una herramienta conveniente y casi indispensable. En este capítulo desarrollaremos y aplicaremos el concepto de vector en dos dimensiones, dejando para el Volumen II las extensiones a dimensiones superiores.

Explicación intuitiva

Muchos entes matemáticos y físicos quedan caracterizados por un solo número, llamado "escalar", ya que mide el ente con relación a una escala. Como ejemplos tenemos ángulos, longitudes, áreas, tiempos, masas y temperaturas. Sin embargo, existen otros entes para los cuales tal caracterización no es posible; por ejemplo, la forma de un triángulo, la posición de un punto en el espacio, la aceleración o la dirección del movimiento de una partícula y el estado de tensión de un cuerpo. Para identificar cada una de estas entidades se requieren varios números. Gradualmente se han ido desarrollando conceptos matemáticos que van más allá del continuo de números reales y que permiten representar tales entidades con un solo símbolo.[1] Los *vectores* en un plano son objetos matemáticos cuya descripción requiere dos componentes de información: una *longitud* y una *dirección.* A este tipo pertenecen, por ejemplo, la posición relativa y orden de dos puntos, la velocidad y la aceleración de una partícula, la fuerza que actúa sobre una partícula, etc.[2]

Geométrica o intuitivamente, un vector es esencialmente un segmento de recta *dirigido,* del plano o del espacio, caracterizado por su longitud o magnitud y por su dirección. Ordinariamente los vectores se indican con *flechas* de la longitud dada y apuntando en la dirección que se especifica. A menos que explícitamente se imponga una restricción, el *vector es "libre",* esto es, la localización del extremo inicial del segmento dirigido (o flecha) no es parte inherente de las especificaciones que se dan para el vector.

A pesar de que conceptos físicos tales como velocidad, aceleración y fuerza son ejemplos primarios de vectores en las aplicaciones, definiremos los vectores geométricamente mediante 'traslaciones" o "desplazamientos paralelos".

El análisis vectorial comienza simplemente dando un nombre, "vector", a tales segmentos de recta dirigidos o desplazamientos paralelos. Sin em-

---

[1] Por supuesto, los números complejos $a + ib = z$ son entes de este tipo, siendo símbolos que representan pares de números reales a, b. Es, sin duda, conveniente algunas veces usar números complejos en vez de vectores.

[2] Los vctores son insuficientes para algunos propósitos; por ejemplo, para describir tensiones o curvaturas de espacios se usan entes más generales llamados "tensores".

bargo, la característica decisiva de los vectores no consiste en tener un nombre que los unifique sino en que (de la misma manera que los números complejos) pueden combinarse entre ellos o con escalares (esto es, números ordinarios) mediante un conjunto de reglas, dando origen al álgebra o análisis vectorial, de manera tal que tienen interpretaciones naturales en las diferentes aplicaciones, como, por ejemplo, la superposición de dos velocidades o el trabajo realizado en un desplazamiento venciendo una fuerza. En el lenguaje vectorial que surge intuitivamente, muchas relaciones matemáticas y físicas pueden expresarse en forma concisa y clara.

a. Definición de vectores mediante traslaciones. Notaciones

El tipo más simple de transformación en el plano es una *traslación* o *desplazamiento paralelo*. Una traslación desplaza o transforma cualquier punto $P = (x, y)$ en el punto $P' = (x', y')$ de coordenadas

$$x' = x + a, \quad y' = y + b,$$

donde a y b son constantes. La traslación está completamente determinada por las constantes a y b, que llamaremos *componentes* de la traslación. Usaremos el término "vector" como otro nombre para "traslación". Usando negritas para denotar vectores o traslaciones, escribimos $\mathbf{R} = (a, b)$ para el vector con componentes a, b (Fig. 4.29).

Las componentes del vector $\mathbf{R}$ están determinadas por un par de puntos correspondientes $P = (x, y)$ y $P' = (x', y')$, ya que entonces

$$a = x' - x, \quad b = y' - y.$$

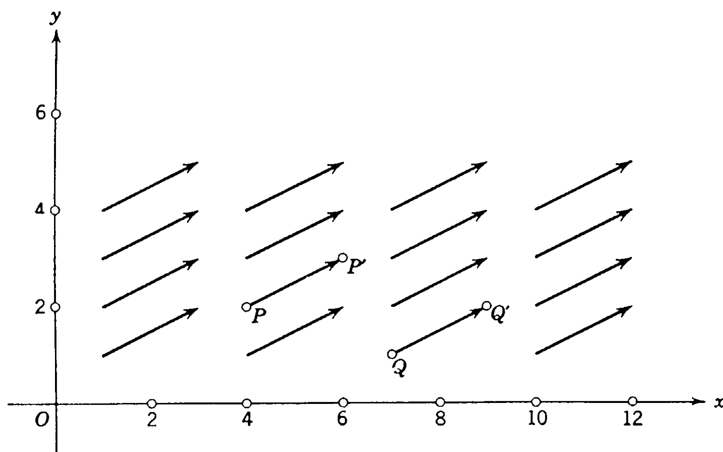


Figura 4.29 La traslación $x' = x + 2$, $y' = y + 1$ correspondiente al vector $\mathbf{R} = \overrightarrow{PP'} = \overrightarrow{QQ'} = (2, 1)$.

Claramente, para puntos P y P' cualesquiera siempre es posible encontrar una traslación **R** que transforme P en P'. A este vector lo denotamos por $\mathbf{R} = \overrightarrow{PP'}$. Cualquier par ordenado de puntos $P = (x, y)$, $P' = (x', y')$, esto es, cualquier segmento orientado determina así el vector $\mathbf{R} = \overrightarrow{PP'} = (x' - x,\ y' = y)$. Obsérvese que un segundo par de punto $Q = (\xi,\ \eta)$, $Q' = (\xi', \eta')$ define el mismo vector si $\xi' - \xi = x' - x$ y $\eta' - \eta = y' - y$; la misma traslación **R** transforma entonces P en P' y Q en Q'. Los vectores **R** están determinados por dos números, sus *componentes*, de la misma manera que los puntos de un plano lo están por sus dos coordenadas. La diferencia básica consiste en que un vector está representado geométricamente por un *par* de puntos. En la representación $\mathbf{R} = \overrightarrow{PP'}$ llamamos a P el *punto inicial* y a P' el *punto final*. Para un **R** dado, uno de los puntos, digamos, el punto inicial $P = (x, y)$, puede escogerse arbitrariamente; entonces el punto final $P' = (x', y')$ queda determinado en forma única por las relaciones $x' = x + a$, $y' = y + b$. Intercambiando los puntos inicial y final se obtiene el vector *opuesto* $\overrightarrow{P'P} = (-a, -b)$.

Si escogemos como punto inicial el origen $O = (0, 0)$, podemos asociar unívocamente un vector **R** con todo punto $Q = (x, y)$ tomando simplemente $\mathbf{R} = \overrightarrow{OQ}$. El vector **R** con punto inicial fijo O se llama entonces *vector de posición* de Q. Las componnentes de la posición del punto Q son simplemente las coordenadas x, y de ese punto.

El vector **R** de componentes $a = 0$, $b = 0$ se llama *vector nulo* y se denota por **O**. Corresponde a una traslación que deja fijo cualquier punto:

$$\mathbf{O} = (0, 0) = \overrightarrow{PP}.$$

La distancia r entre dos puntos $P = (x, y)$ y $P' = (x', y')$ después sólo del vector $\mathbf{R} = (a, b) = \overrightarrow{PP'}$, puesto que

$$r = \sqrt{(x' - x)^2 + (y' - y)^2} = \sqrt{a^2 + b^2}.$$

A esta distancia le llamamos la *longitud del vector* **R**, y escribimos $r = \mathbf{R}$. La longitud de **R** es siempre un número positivo a menos que $\mathbf{R} = \mathbf{O}$ (ver Fig. 4.30).

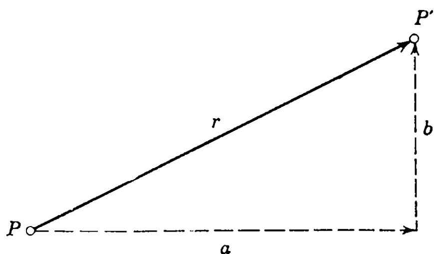


Figura 4.30 Componentes a, b y longitud r de un vector $\mathbf{R} = \overrightarrow{PP'}$.

Definimos el *producto* de un vector $\mathbf{R} = (a, b)$ por un número o "escalar" λ como el vector

$$\mathbf{R}^* = \lambda\mathbf{R} = (\lambda a, \lambda b).$$

Con $\lambda = -1$ tenemos en $\mathbf{R}^* = (-a, -b)$ el vector *opuesto* a $\mathbf{R}$ (Fig. 4.31).

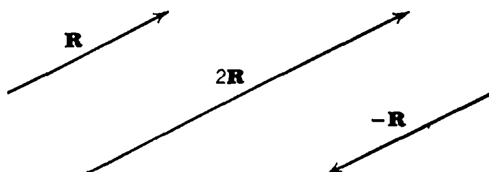


Figura 4.31 Múltiplos escalares de un vector R.

Si $\mathbf{R} = \overrightarrow{PP'} = (a, b)$, con $P = (x, y)$, $P' = (x', y')$, podemos representar $\mathbf{R}^* = \lambda\mathbf{R}$ como $\overrightarrow{PP''}$, donde $P'' = (x'', y'') = (x + \lambda a, y + \lambda b)$ (ver Fig. 4.32). Para $a = b = 0$ se tiene, por supuesto, $P'' = P' = P$. Para a y b no simultáneamente nulos, el punto $P'' = (x'', y'') = (x + \lambda a, y + \lambda b)$ recorre completamente, para λ variable, la recta

$$x''b - y''a = xb - ya.$$

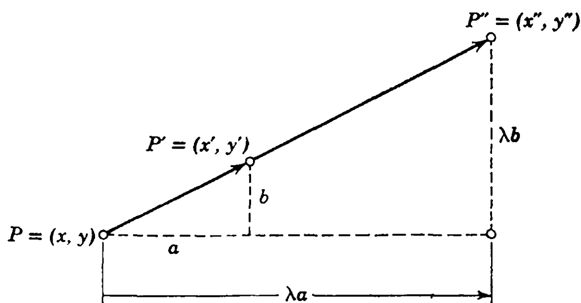


Figura 4.32 La relación vectorial $\mathbf{R}^* = \overrightarrow{PP''} = \lambda\,\overrightarrow{PP'}$ para $\lambda = 8/3$.

El valor $\lambda = 0$ da $P'' = P$, en tanto que $\lambda = 1$ da $P'' = P'$. Así, P'' está sobre la recta que pasa por P y P'. Para $\lambda > 0$ los puntos P'' y P' están sobre el *mismo lado* de P; para $\lambda < 0$, están en *lados opuestos.*

Los dos vectores $\mathbf{R} = (a, b)$ y $\mathbf{R}^* = (a^*, b^*)$ se dice que tienen el *mismo sentido* si $\mathbf{R}^* = \lambda\mathbf{R}$ con λ positivo, y sentidos opuestos si $\lambda < 0$. Si $\mathbf{R} = \mathbf{O}$, esto significa que también $\mathbf{R}^* = \mathbf{O}$. Si $\mathbf{R} \neq \mathbf{O}$, la condición necesaria y suficiente para que $\mathbf{R}^*$ tenga el mismo sentido que $\mathbf{R}$ es que

$$\frac{a}{\sqrt{a^2+b^2}} = \frac{a^*}{\sqrt{a^{*2}+b^{*2}}}, \qquad \frac{b}{\sqrt{a^2+b^2}} = \frac{b^*}{\sqrt{a^{*2}+b^{*2}}}.$$

A las cantidades

$$\xi = \frac{a}{\sqrt{a^2+b^2}} = \frac{a}{|\mathbf{R}|} = \frac{a}{r}, \qquad \eta = \frac{b}{\sqrt{a^2+b^2}} = \frac{b}{|\mathbf{R}|} = \frac{b}{r},$$

que determinan la dirección del vector, $\mathbf{R}$ les llamamos *cosenos directores* de $\mathbf{R}$. No están definidos, por supuesto, para $\mathbf{R} = \mathbf{O}$. Puesto que $\xi^2 + \eta^2 = 1$, podemos siempre encontrar un ángulo α y un correspondiente ángulo $\beta = \pi/2 - \alpha$ tales que

$$\xi = \cos\alpha, \qquad \eta = \operatorname{sen}\alpha = \cos\beta.$$

El ángulo α se llama *ángulo de dirección* de $\mathbf{R}$ (Fig. 4.33). Está determinado sólo a menos de un múltiplo par de π. Para $\mathbf{R} = \overrightarrow{PP'}$ tenemos

$$\cos\alpha = \frac{x'-x}{r}, \qquad \operatorname{sen}\alpha = \frac{y'-y}{r}.$$

Obviamente, α es el ángulo entre el eje x positivo y la recta que va de P a P'. Más precisamente, una rotación del eje x positivo alrededor del origen en un ángulo α (considerado positivo si la rotación es contra reloj, negativo en el caso opuesto) dará a dicho eje la dirección de P a P'. El vector opuesto $-\mathbf{R} = (-a, -b)$ tiene cosenos directores $-\xi$ y $-\eta$, y ángulos de dirección que difieren de α en múltiplos impares de π. Si el punto inicial P del vector $\mathbf{R} = \overrightarrow{PP'}$ es el origen, el ángulo de dirección α de $\mathbf{R}$ es simplemente el ángulo polar θ de P'.

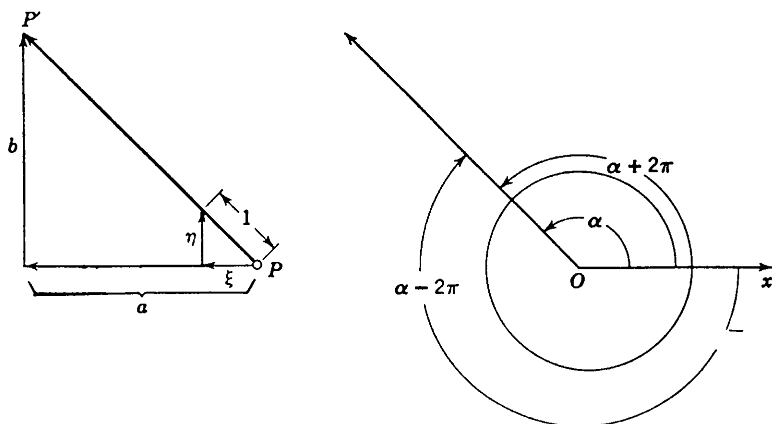


Figura 4.33 Cosenos directores ξ, η y ángulos de dirección para un vector $\overrightarrow{PP'}$.

b. Adición y multiplicación de vectores

Suma de vectores

Los vectores se han definido mediante traslaciones, esto es, como ciertas transformaciones de puntos en el plano. Existe una manera perfectamente general de combinar dos transformaciones cualesquiera aplicándolas *sucesivamente*.

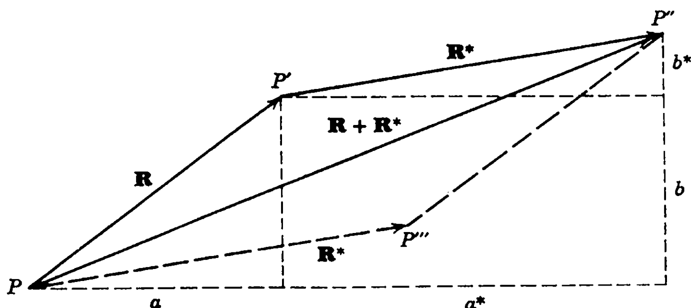


Figura 4.34 Adición de los vectores $\overrightarrow{PP'} = (a, b)$ y $\overrightarrow{P'P''} = (a^*, b^*)$.

Si la primera transformación lleva el punto P al punto P' y la segunda lleva P' a P'', la transformación combinada es aquélla que lleva P a P''. En el caso de dos vectores $\mathbf{R} = (a, b)$ y $\mathbf{R}^* = (a^*, b^*)$, el vector $\mathbf{R}$ transformará el punto $P = (x, y)$ en el punto $P' = (x + a, y + b)$ y $\mathbf{R}^*$ transformará P' en $P'' = (x + a + a^*, y + b + b^*)$. La transformación de P en P'' resultante es también una traslación; la llamaremos la *suma* o la *resultante de los vectores* $\mathbf{R} = \overrightarrow{PP'}$ y $\mathbf{R}^* = \overrightarrow{P'P''}$, y la denotaremos por $\mathbf{R} + \mathbf{R}^*$ (Fig. 4.34).[1] Las componentes de la suma son $a + a^*$ y $b + b^*$. Así, nuestra definición de suma de dos vectores es

$$\overrightarrow{PP'} + \overrightarrow{P'P''} = \overrightarrow{PP''},$$

o bien, si describimos los vectores por sus componentes,

$$(a, b) + (a^*, b^*) = (a + a^*, b + b^*).$$

Si $\mathbf{R}^*$ se toma con el mismo punto inicial que $\mathbf{R}$, digamos $\mathbf{R}^* = \overrightarrow{PP'''}$, los puntos P, P''', P'' y P' constituyen los vértices de un paralelogramo. Los dos lados que parten de P representan los vectores $\mathbf{R}$ y $\mathbf{R}^*$. La suma $\mathbf{R} + \mathbf{R}^*$ está representada por la *diagonal* que parte de P (*"construcción del paralelogramo" para la suma de vectores*).

[1] Esta "suma" es realmente el "producto símbólico" de las dos transformaciones, como se define en la p. 76. La notación de suma es aquí más natural debido a que corresponde a la adición de las componentes.

Las sumas vectoriales obedecen las leyes conmutativa y asociativa de la aritmética, ya que la adición de vectores se reduce a la adición de las componentes correspondientes (Fig. 4.35). Dichas sumas obedecen además las *leyes distributivas para la multiplicación de una suma de dos vectores por un número* λ *y de un vector por la suma de dos números* λ, μ:

$$\lambda(\mathbf{R} + \mathbf{R}^*) = \lambda\mathbf{R} + \lambda\mathbf{R}^*, \qquad (\lambda + \mu)\mathbf{R} = \lambda\mathbf{R} + \mu\mathbf{R}.$$ [2]

Estas reglas nos permiten expresar un vector $\overrightarrow{PP'}$ en términos de los vectores de posición $\overrightarrow{OP}$ y $\overrightarrow{OP'}$ de los puntos P y P' (Fig. 4.36):

$$\overrightarrow{PP'} = \overrightarrow{PO} + \overrightarrow{OP'} = \overrightarrow{OP'} + \overrightarrow{PO} = \overrightarrow{OP'} - \overrightarrow{OP}.$$

Es importante darse cuenta de que, generalmente, si se va de un punto P a un punto Q pasando por los puntos $A, B, C, \ldots, E, F$, entonces el vector $\overrightarrow{PQ}$ es la suma de los vectores $\overrightarrow{PA}, \overrightarrow{AB}, \overrightarrow{BC}, \ldots, \overrightarrow{EF}, \overrightarrow{FQ}$ (Fig. 4.37).

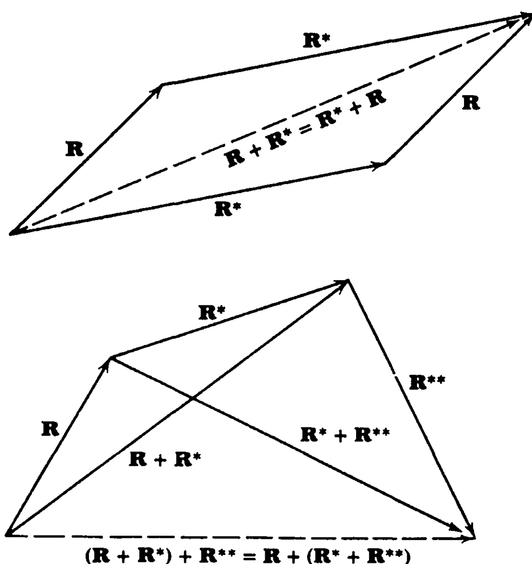


Figura 4.35 Leyes conmutativa y asociativa de la adición vectorial.

Angulo entre vectores

El ángulo θ formado por un vector $\mathbf{R}^* = (a^*, b^*)$ con el vector $\mathbf{R} = (a, b)$ se define como la diferencia entre sus ángulos de dirección:

[2] Para distinguir entre vectores y números en una ecuación, siempre ponemos el número antes del vector al escribir productos. La combinación $\mathbf{R}\lambda$ no se usará, aunque podría definirse por $\lambda\mathbf{R} = \mathbf{R}\lambda$.

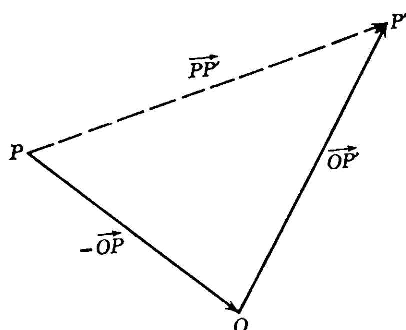


Figura 4.36 $\overrightarrow{PP'} = \overrightarrow{OP'} - \overrightarrow{OP}$.

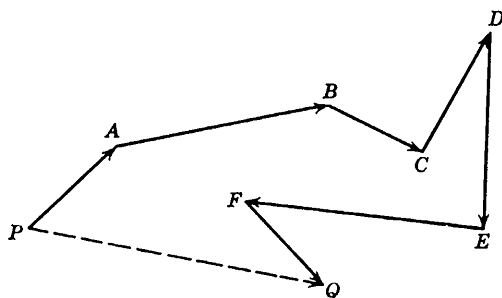


Figura 4.37 $\overrightarrow{PQ} = \overrightarrow{PA} + \overrightarrow{AB} + \overrightarrow{BC} + \cdots + \overrightarrow{FQ}$.

$\theta = \alpha^* - \alpha$. (Aquí se supone que ninguno de los vectores **R** y **R*** es el vector cero.) Nuevamente, el ángulo θ está determinado solamente a menos de múltiplos enteros de 2π (Fig. 4.38). Una rotación de ángulo θ (con el signo de θ indicando el sentido de la rotación) transformará la dirección de **R** en la de **R***. Las cantidades $\cos\theta$ y $\operatorname{sen}\theta$, que están determinadas unívocamente, pueden inmediatamente expresarse en términos de los cosenos directores de **R** y **R***:

$$\cos\theta = \cos(\alpha^* - \alpha) = \cos\alpha\cos\alpha^* + \operatorname{sen}\alpha\operatorname{sen}\alpha^*$$

$$= \frac{aa^* + bb^*}{\sqrt{a^2 + b^2}\sqrt{a^{*2} + b^{*2}}}$$

$$\operatorname{sen}\theta = \operatorname{sen}(\alpha^* - \alpha) = \cos\alpha\operatorname{sen}\alpha^* - \operatorname{sen}\alpha\cos\alpha^*$$

$$= \frac{ab^* - a^*b}{\sqrt{a^2 + b^2}\sqrt{a^{*2} + b^{*2}}}$$

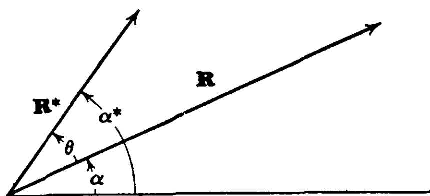


Figura 4.38. Angulo θ que el vector $\mathbf{R}^*$ forma con $\mathbf{R}$.

El denominador de cada expresión es justamente el producto rr^* de las longitudes de los vectores. Las expresiones que aparecen en los numeradores se introducen como "productos" de los dos vectores.

Producto interior y producto exterior de dos vectores

Se define el producto "escalar", producto 'interno" o producto "punto" de dos vectores, $\mathbf{R} = (a, b)$ y $\mathbf{R}^* = (a^*, b^*)$, mediante la expresión

$$\mathbf{R} \cdot \mathbf{R}^* = aa^* + bb^* = rr^* \cos \theta,$$

y el producto "exterior" o producto "cruz", por la expresión

$$\mathbf{R} \times \mathbf{R}^* = ab^* - a^*b = rr^* \operatorname{sen} \theta.$$[1]

Como puede fácilmente verificarse, *los productos interior y exterior obedecen las leyes distributiva y asociativa:*

$$\mathbf{R} \cdot (\mathbf{R}^* + \mathbf{R}^{**}) = \mathbf{R} \cdot \mathbf{R}^* + \mathbf{R} \cdot \mathbf{R}^{**},$$

$$\mathbf{R} \times (\mathbf{R}^* + \mathbf{R}^{**}) = \mathbf{R} \times \mathbf{R}^* + \mathbf{R} \times \mathbf{R}^{**},$$

$$\lambda(\mathbf{R} \cdot \mathbf{R}^*) = (\lambda\mathbf{R}) \cdot \mathbf{R}^* = \mathbf{R} \cdot (\lambda\mathbf{R}^*),$$

$$\lambda(\mathbf{R} \times \mathbf{R}^*) = (\lambda\mathbf{R}) \times \mathbf{R}^* = \mathbf{R} \times (\lambda\mathbf{R}^*).$$

La ley conmutativa de la multiplicación también se cumple para productos interiores:

$$\mathbf{R} \cdot \mathbf{R}^* = \mathbf{R}^* \cdot \mathbf{R}.$$

Sin embargo, para productos exteriores el intercambio de los factores cambia el signo:

$$\mathbf{R} \times \mathbf{R}^* = -\mathbf{R}^* \times \mathbf{R}.$$

Dados $\mathbf{R}$ y $\mathbf{R}^*$ con el mismo punto inicial, $\mathbf{R} = \overrightarrow{PQ}$, $\mathbf{R}^* = \overrightarrow{PQ}$, podemos interpretar $\mathbf{R} \cdot \mathbf{R}^*$ como el producto de la proyección $r^* \cos \theta$ del seg-

[1] Con nuestra definición, tanto el producto interior como el exterior son realmente "escalares". El término "producto escalar" se reserva para el producto interior debido a que en tres dimensiones el análogo del producto exterior es un vector (en este caso, el producto exterior también se llama *"producto vectorial")*.

mento PQ^* sobre el segmento PQ y la longitud r de este segmento. El producto exterior $\mathbf{R} \times \mathbf{R}^*$ es simplemente el doble del área del triángulo orientado PQQ^*, tomada con signo positivo si los vértices PQQ^* están ordenados contra reloj y con signo negativo para el orden inverso (Fig. 4.39).

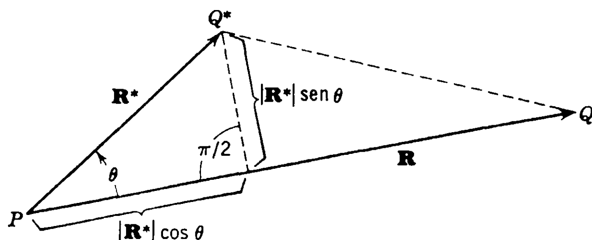


Figura 4.39 El producto vectorial $\mathbf{R} \times \mathbf{R}^* = |\mathbf{R}|\,|\mathbf{R}^*| \operatorname{sen} \theta$ como el doble del área del triángulo PQQ^*.

Para cualquier vector $\mathbf{R} = (a, b)$,

$$\mathbf{R} \cdot \mathbf{R} = a^2 + b^2 = |\mathbf{R}|^2$$

es el cuadrado de la longitud del vector. Así, $\mathbf{R} \cdot \mathbf{R}$ es positivo a menos que $\mathbf{R} = \mathbf{O}$. Por otra parte, $\mathbf{R} \times \mathbf{R}$ es siempre cero. La condición para que dos vectores no nulos sean mutuamente ortogonales es que $\mathbf{R} \cdot \mathbf{R}^* = 0$, y serán paralelos (esto es, tendrán la misma dirección y sentidos iguales u opuestos) si $\mathbf{R} \times \mathbf{R}^* = 0$.

Ecuación de la recta

Usando la notación vectorial podemos escribir fácilmente la ecuación de una recta que pasa por dos puntos. Si $P = (x, y)$, $P_0 = (x_0, y_0)$ y $P_1 = (x_1, y_1)$ son tres puntos tales que $P_0 \neq P_1$, entonces P estará sobre la recta que pasa por P_0 y P_1 si los vectores $\overrightarrow{P_0P}$ y $\overrightarrow{P_0P_1}$ son paralelos, esto es, si

$$\overrightarrow{P_0P} \times \overrightarrow{P_0P_1} = \mathbf{0}.$$

Si $\mathbf{R} = \overrightarrow{OP}$, $\mathbf{R}_0 = \overrightarrow{OP_0}$ y $\mathbf{R}_1 = \overrightarrow{OP_1}$ son los vectores de posición de los tres puntos, la condición toma la forma

$$(\mathbf{R} - \mathbf{R}_0) \times (\mathbf{R}_1 - \mathbf{R}_0) = \mathbf{0}$$

o bien

$$(\mathbf{R}_1 - \mathbf{R}_0) \times \mathbf{R} = \mathbf{R}_1 \times \mathbf{R}_0.$$

Substituyendo los vectores de posición por las coordenadas de los puntos se obtiene la ecuación de la recta en la forma usual (Fig. 4.40):

$$(x_1 - x_0)y - (y_1 - y_0)x = x_1y_0 - y_1x_0.$$

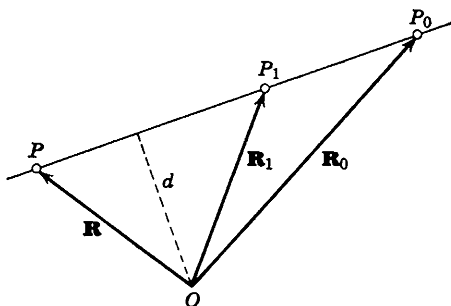


Figura 4.40 La recta en notación vectorial.

En vez de especificar dos puntos de la recta, podemos especificar sólo uno, P_0, y requerir que la recta sea paralela al vector $\mathbf{S} = (a, b)$. Evidentemente, la ecuación de la recta es entonces

$$(\mathbf{R} - \mathbf{R}_0) \times \mathbf{S} = 0,$$

o bien

$$(x - x_0)b - (y - y_0)a = 0.$$

Para $\mathbf{S} = \overrightarrow{P_0P_1}$ se obtiene la ecuación primera.

La distancia d de la recta al origen puede también expresarse en notación vectorial. Es claro que el producto de d y la longitud del vector $\overrightarrow{P_0P_1}$ es el doble del área del triángulo OP_0P_1. Por lo tanto,

$$d = \frac{1}{|P_0P_1|}(\overrightarrow{OP_0} \times \overrightarrow{OP_1}) = \frac{\mathbf{R}_0 \times \mathbf{R}_1}{|\mathbf{R}_1 - \mathbf{R}_0|}$$
$$= \frac{x_0y_1 - x_1y_0}{\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}}.$$

Aquí se toma d con signo positivo si los puntos O, P_0 y P_1 se suceden uno al otro en el sentido contra reloj.

Vectores coordenados. Un vector $\mathbf{R} = (a, b)$ puede representarse en la forma

(29) $$\mathbf{R} = a\mathbf{i} + b\mathbf{j},$$

donde se denotan por $\mathbf{i}$ y $\mathbf{j}$ los *"vectores coordenados"*

(30) $$\mathbf{i} = (1, 0), \qquad \mathbf{j} = (0, 1).$$

De esta manera se divide $\mathbf{R}$ en dos vectores $a\mathbf{i}$ y $b\mathbf{j}$ que apuntan respectivamente en las direcciones positivas de los ejes x e y. Las componentes a y b de $\mathbf{R}$ son justamente las longitudes (con sus signos) de estos dos vectores.

En las aplicaciones se tiene a menudo necesidad de representar un vector **R** como la resultante de vectores con dos direcciones ortogonales (esto es, mutuamente perpendiculares) dadas. Con este fin, lo mejor es introducir dos *vectores unitarios* (o sea, vectores de longitud 1) **I** y **J** con las direcciones prescritas. La descomposición de **R** requerida se realiza si puede representarse **R** en la forma

$$\mathbf{R} = A\mathbf{I} + B\mathbf{J} \tag{31}$$

con escalares adecuados A, B (cfr. Fig. 4.40). Es fácil encontrar los valores de A y B en el caso en que tal representación de **R** exista. Pues, en efecto, por hipótesis, los vectores **I** y **J** son vectores unitarios de longitud 1, de manera que

$$\mathbf{I} \cdot \mathbf{I} = \mathbf{J} \cdot \mathbf{J} = 1, \qquad \mathbf{I} \cdot \mathbf{J} = 0. \tag{32}$$

Formando el producto escalar de la Ec. (31) con **I**, **J**, respectivamente, se encuentra de inmediato que A y B deben tener los valores

$$A = \mathbf{R} \cdot \mathbf{I}, \qquad B = \mathbf{R} \cdot \mathbf{J}; \tag{33}$$

es decir, A y B son las longitudes (con sus signos respectivos) de las proyecciones en las direcciones dadas del segmento que representa a **R**.

La posibilidad de expresar **R** como una combinación lineal (31) de **I** y **J** se deriva de la representación (29) de **R** en términos de **i** y **j**, siempre que pueda demostrarse que **i** y **j** mismos pueden ser expresados en términos de **I** y **J**. Sin embargo, $\mathbf{I} = (\alpha, \beta)$ y $\mathbf{J} = (\gamma, \delta)$ pueden escribirse de la manera siguiente

$$\mathbf{I} = \alpha\mathbf{i} + \beta\mathbf{j}, \qquad \mathbf{J} = \gamma\mathbf{i} + \delta\mathbf{j}. \tag{34}$$

Como consecuencia de (32), las cantidades α, β, γ, δ deben satisfacer las llamadas relaciones de ortogonalidad

$$\alpha^2 + \beta^2 = \gamma^2 + \delta^2 = 1, \qquad \alpha\gamma + \beta\delta = 0. \tag{35}$$

Si se multiplica la primera de las ecuaciones (34) por δ, la segunda por β, y se restan, se obtiene

$$(\alpha\delta - \beta\gamma)\mathbf{i} = \delta\mathbf{I} - \beta\mathbf{J}; \tag{36}$$

y, en forma análoga,

$$(\alpha\delta - \beta\gamma)\mathbf{j} = -\gamma\mathbf{I} + \alpha\mathbf{J}. \tag{37}$$

Aquí se tiene para los vectores unitarios mutuamente perpendiculares **I** y **J** que

$$\alpha\delta - \beta\gamma = \mathbf{I} \times \mathbf{J} = \pm 1, \tag{38}$$

donde se escoge un signo u otro dependiendo de que el sentido de la rotación en 90° que transforme **I** en **J** sea contra reloj o en el sentido del reloj. En uno u otro caso las fórmulas (36) y (37) expresan **i** y **j** en términos de **I** y **J**. Substituyendo estas expresiones en (29) se obtiene la justificación de la fórmula de representación (31) para un vector arbitrario **R**.

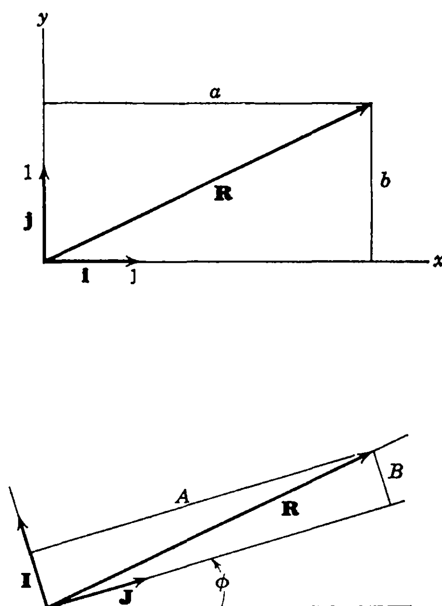


Figura 4.40

La fórmula (31) puede también interpretarse como la representación del vector **R** en un nuevo sistema de coordenadas cuyos ejes apuntan respectivamente en las direcciones de **I** y **J**. Las componentes de un vector unitario son al mismo tiempo los cosenos directores del ángulo de dirección ϕ y ψ, respectivamente. Entonces,

$$\alpha = \cos\phi, \qquad \beta = \operatorname{sen}\phi, \qquad \gamma = \cos\psi, \qquad \delta = \operatorname{sen}\psi.$$

Aquí se tiene $\psi = \phi + \frac{1}{2}\pi$, o bien $\psi = \phi - \frac{1}{2}\pi$. En el primer caso (que corresponde a un sistema dextrógiro con vectores coordenados **I** y **J**) se tiene $\gamma = -\beta$, $\delta = \alpha$, $\alpha\delta - \beta\gamma = +1$; de manera que

(39) $$\mathbf{I} = (\cos\phi, \operatorname{sen}\phi), \qquad \mathbf{J} = (-\operatorname{sen}\phi, \cos\phi).$$

Las fórmulas (33) que dan las componentes de **R** referidas a los vectores coordenados **I** y **J** toman entonces la forma

(40) $$A = a\cos\phi + b\operatorname{sen}\phi, \qquad B = -a\operatorname{sen}\phi + b\cos\phi.$$

Estas fórmulas expresan las relaciones entre las componentes de uno y el mismo vector **R** en dos sistemas de coordenadas dextrógiros que se obtienen uno del otro mediante una rotación de ejes en un ángulo ϕ. Si se supone que los sistemas de coordenadas tienen el mismo origen O y que **R** es el vector de posición $\overrightarrow{OP}$ de un punto arbitrario P, las fórmulas (40) expresan la transformación de los sistemas de coordenadas obtenidas ya en la p. 379, ecuación (18). Las componentes a, b y A, B son entonces respectivamente las coordenadas de P en los dos sistemas.

c. Vectores variables, sus derivadas e integrales

Es natural considerar vectores $\mathbf{R} = (a, b)$ cuyas componentes a, b sean funciones de una variable t, digamos $a = a(t)$, $b = b(t)$. Para cualquier t se tiene entonces un vector

$$\mathbf{R} = \mathbf{R}(t) = (a(t), b(t))$$

y se dice que $\mathbf{R}(t)$ es una *función vectorial* de t. Un ejemplo lo proporciona el vector de posición de un punto que se mueve al variar el tiempo t.

Se dice que $\mathbf{R}(t)$ tiene el límite $\mathbf{R}^* = (a^*, b^*)$ para $t \to t_0$ si $a(t)$ tiene el límite a^* y $b(t)$ el límite b^* para $t \to t_0$. En ese caso la longitud de $\mathbf{R}(t)$ tiende hacia la de $\mathbf{R}^*$ y, si $\mathbf{R}^* \neq 0$, la dirección de $\mathbf{R}(t)$ tiende hacia la de $\mathbf{R}^*$ (esto significa que los cosenos directores de **R** tienden hacia los de $\mathbf{R}^*$). Se dice que el vector $\mathbf{R}(t)$ depende continuamente de t si

$$\lim_{t \to t_0} \mathbf{R}(t) = \mathbf{R}(t_0),$$

esto es, si las componentes de **R** son funciones continuas de t. La longitud y, si $\mathbf{R}(t_0) \neq \mathbf{O}$, también la dirección de un vector continuo varían continuamente con t.

Para introducir la *derivada de un vector* se forma para dos valores t y $t + h$ del parámetro el cociente incremental

$$\frac{1}{h}[\mathbf{R}(t+h) - \mathbf{R}(t)] = \left[\frac{a(t+h) - a(t)}{h}, \frac{b(t+h) - b(t)}{h}\right],$$

y se define la derivada de **R** como el límite del cociente incremental para $h \to 0$:

$$\dot{\mathbf{R}} = \frac{d\mathbf{R}}{dt} = \lim_{h \to 0} \frac{1}{h}[\mathbf{R}(t+h) - \mathbf{R}(t)] = \left(\frac{da}{dt}, \frac{db}{dt}\right) = (\dot{a}, \dot{b}).$$

La derivada de un vector se forma derivando sus componentes.

Las derivadas de productos de vectores, como se ve fácilmente, obedecen las reglas ordinarias de las derivadas de productos de funciones reales o "escalares":

$$(\mathbf{RS})^{\cdot} = \frac{d\mathbf{R}\cdot\mathbf{S}}{dt} = \frac{d\mathbf{R}}{dt}\cdot\mathbf{S} + \mathbf{R}\cdot\frac{d\mathbf{S}}{dt} = \dot{\mathbf{R}}\mathbf{S} + \mathbf{R}\dot{\mathbf{S}}$$

$$(\mathbf{R}\times\mathbf{S})^{\cdot} = \frac{d\mathbf{R}\times\mathbf{S}}{dt} = \dot{\mathbf{R}}\times\mathbf{S} + \mathbf{R}\times\dot{\mathbf{S}},$$

donde para los productos exteriores es necesario conservar el orden original de los factores.

En forma análoga se define la integral del vector $\mathbf{R}(t)$ en términos de las integrales de sus componentes:

$$\int_{\alpha}^{\beta}\mathbf{R}(t)\,dt = \left(\int_{\alpha}^{\beta}a(t)\,dt, \int_{\alpha}^{\beta}b(t)\,dt\right).$$

El teorema fundamental del cálculo implica que

$$\frac{d}{dt}\int_{\alpha}^{t}\mathbf{R}(s)\,ds = \mathbf{R}(t).$$

d. Aplicación a curvas planas. Dirección, "rapidez" y aceleración

Vector velocidad

En la Sección 4.1 se representó una curva C mediante dos funciones $x = \phi(t)$ e $y = \psi(t)$. Cada valor de t en el dominio de estas funciones determina un punto $P = (x, y)$ sobre C; en este caso t puede considerarse como el tiempo y P como un punto móvil cuya posición en el tiempo t está dada por $x(t)$ e $y(t)$. Si se identifican x e y con las componentes del vector de posición $\mathbf{R} = \overrightarrow{OP}$ de P, entonces C queda descripta por el punto final del vector de posición

$$\mathbf{R} = \mathbf{R}(t) = (x(t), y(t))$$

(Fig. 4.41). Para dos puntos P y P' de C correspondientes a los valores paramétricos t y $t + \Delta t$, la expresión

$$\overrightarrow{PP'} = \overrightarrow{OP'} - \overrightarrow{OP} = \mathbf{R}(t+\Delta t) - \mathbf{R}(t) = \Delta\mathbf{R}$$

da el vector representado por la secante dirigida de C con puntos extremos P, P'. Aquí Δt es positivo, esto es, si el punto P' sigue a P sobre C en el sentido de las t crecientes, entonces el vector

$$\frac{t}{\Delta t}(\mathbf{R}(t+\Delta t) - \mathbf{R}(t))$$

tiene la misma dirección que el vector $\mathbf{R}(t + \Delta t) - \mathbf{R}(t) = \overrightarrow{\mathrm{PP}'}$; su longitud es la distancia entre los puntos P y P' dividida por Δt. Cuando Δt tiende a cero se obtiene en el límite el vector

$$\dot{\mathbf{R}} = \dot{\mathbf{R}}(t) = (\dot{x}(t), \dot{y}(t)),$$

donde, nuevamente, el punto se usa para denotar derivación con respecto al parámetro t. La dirección de $\dot{\mathbf{R}}$ es el límite de las direcciones de las secantes PP', y, por lo tanto, es la dirección de la tangente en el punto P. Más precisamente, $\dot{\mathbf{R}}$ apunta en aquella dirección de la tangente que co-

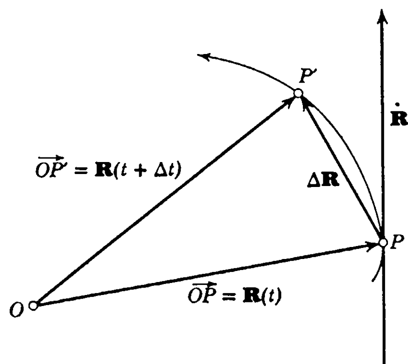


Figura 4.41 Derivada del vector de posición para una curva.

rresponde a valores crecientes de t sobre C, suponiendo que $\dot{\mathbf{R}} \neq \mathbf{O}$. Los cosenos directores de $\dot{\mathbf{R}}$ son las cantidades

$$\cos\alpha = \frac{\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2}}, \qquad \operatorname{sen}\alpha = \frac{\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}},$$

introducidas en la p. 364 como cosenos directores de la tangente. La longitud de $\dot{\mathbf{R}}$,

$$|\dot{\mathbf{R}}| = \sqrt{\dot{x}^2 + \dot{y}^2},$$

puede interpretarse como ds/dt, esto es, la medida de cambio, con respecto al parámetro t, de la longitud s a lo largo de la curva. Si t representa el tiempo, $|\dot{\mathbf{R}}|$ representa entonces la velocidad con la que el punto se mueve a lo largo de la curva.

En Mecánica uno debe considerar la velocidad de una partícula no sólo como algo que tiene una cierta magnitud (la "rapidez") sino también una cierta *dirección*. La *velocidad* está entonces representada por el *vector* $\dot{\mathbf{R}} = (\dot{x}, \dot{y})$, cuya longitud es la velocidad (escalar) y cuya dirección es la dirección instantánea de movimiento, esto es, la dirección de la tangente en el sentido de las t crecientes.

Aceleración

Análogamente, la *aceleración* de la partícula se define como el vector $\ddot{\mathbf{R}} = (\ddot{x}, \ddot{y})$. Si la aceleración se anula, esto significa que $\ddot{x} = \ddot{y} = 0$; si $\ddot{\mathbf{R}} = \mathbf{O}$ a lo largo de un intervalo t completo, las componentes de la velocidad tienen valores constantes $\dot{x} = a$, $\dot{y} = b$; y las componentes del vector de posición mismo son entonces funciones lineales de t: $x = at + c$, $y = bt + d$. En este caso la partícula se mueve con velocidad constante a lo largo de una línea recta.

Todos nuestros resultados anteriores relativos a curvas son fácilmente expresables en notación vectorial si se describe la curva por el vector de posición $\mathbf{R} = \mathbf{R}(t) = (x(t), y(t))$, con $\alpha \leq t \leq \beta$. Se encuentra para la longitud [cfr. Ec. (8), p. 367] la expresión

$$\int_\alpha^\beta |\dot{\mathbf{R}}|\, dt,$$

en tanto que para que el área, con su signo, encerrada por una curva [cfr. Ec. (20), p. 383]:

$$A = \tfrac{1}{2} \int_\alpha^\beta \mathbf{R} \times \dot{\mathbf{R}}\, dt$$

(el signo de esta cantidad depende de la orientación de la curva). Finalmente, para la curva κ se tiene la fórmula [cfr. Ec. (15), p. 373]

$$\kappa = \frac{\dot{\mathbf{R}} \times \ddot{\mathbf{R}}}{|\dot{\mathbf{R}}|^3}.$$

Componentes tangencial y normal de la aceleración

Las fórmulas anteriores tienen implicaciones interesantes si nuevamente se interpreta t como el tiempo. Sea γ el ángulo formado por el vector $\ddot{\mathbf{R}}$ con el vector $\dot{\mathbf{R}}$, esto es, con la dirección instantánea de movimiento. La cantidad $|\ddot{\mathbf{R}}| \cos\gamma$ representa la proyección de $\ddot{\mathbf{R}}$ en la dirección de $\dot{\mathbf{R}}$; la llamamos *componente tangencial* de la aceleración. Análogamente, $|\ddot{\mathbf{R}}| \operatorname{sen}\gamma$ es la proyección de $\ddot{\mathbf{R}}$ sobre la normal (más precisamente, sobre la normal que se obtiene a partir de $\dot{\mathbf{R}}$ mediante una rotación contra reloj de 90°); ésta es la *componente normal* de la aceleración (ver Fig. 4.42). Por definición de productos interior y exterior,

$$|\ddot{\mathbf{R}}| \cos\gamma = \frac{\dot{\mathbf{R}} \cdot \ddot{\mathbf{R}}}{|\dot{\mathbf{R}}|} \qquad |\ddot{\mathbf{R}}| \operatorname{sen}\gamma = \frac{\dot{\mathbf{R}} \times \ddot{\mathbf{R}}}{|\dot{\mathbf{R}}|}.$$

Ahora bien,

$$\dot{\mathbf{R}} \cdot \ddot{\mathbf{R}} = \frac{1}{2}(\dot{\mathbf{R}} \cdot \ddot{\mathbf{R}} + \ddot{\mathbf{R}} \cdot \dot{\mathbf{R}}) = \frac{1}{2}\frac{d}{dt}(\dot{\mathbf{R}} \cdot \dot{\mathbf{R}}) = \frac{1}{2}\frac{dv^2}{dt} = v\frac{dv}{dt},$$

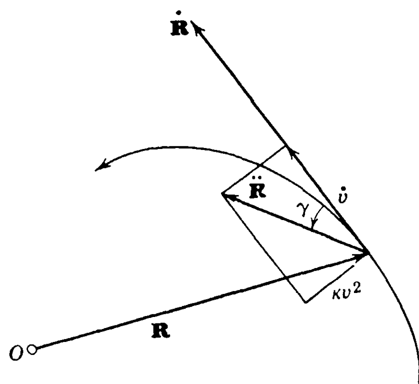


Figura 4.42 Aceleración tangencial y aceleración normal.

donde $v = ds/dt = |\dot{\mathbf{R}}| = \sqrt{\dot{\mathbf{R}} \cdot \dot{\mathbf{R}}}$ es la *velocidad* (escalar) del punto. Por lo tanto

$$(41) \qquad |\ddot{\mathbf{R}}| \cos \gamma = \frac{dv}{dt} = \dot{v};$$

De esta manera, la componente tangencial de la aceleración es idéntica a la medida de cambio, o sea, la derivada, con respecto al tiempo, de la velocidad. Para la aceleración normal la fórmula de la curvatura conduce a

$$(42) \qquad |\ddot{\mathbf{R}}| \operatorname{sen} \gamma = \kappa\, |\dot{\mathbf{R}}|^2 = \kappa v^2,$$

esto es, el producto del cuadrado de la velocidad por la curvatura.

Para una partícula que se mueve con velocidad *constante* v a lo largo de una curva, la aceleración tangencial $\dot{v}$ se anula. El vector aceleración es entonces perpendicular a la curva. Más precisamente, dicho vector apunta hacia el lado "interno" de la curva, o sea, el lado hacia el cual la curva se dobla (esto se ve, por ejemplo, del hecho de que $\operatorname{sen} \gamma > 0$ cuando $\kappa > 0$, o, lo que es lo mismo, cuando la tangente gira contra reloj). Al moverse a lo largo de una curva con rapidez constante un punto experimenta, por lo tanto, una aceleración hacia el interior de la curva, que es proporcional a la curvatura y también al cuadrado de la velocidad. Este hecho es de obvia significación debido a que, como resultado de la ley de Newton (discutida más adelante), se necesita una fuerza proporcional a la aceleración para mantener el punto P sobre la curva.

4.4 Movimiento de una partícula bajo la acción de fuerzas especificadas

Los primeros desarrollos del cálculo fueron estimulados en forma decisiva no solamente por la geometría sino también en una gran medida por

los conceptos de la Mecánica. Esta ciencia descansa sobre ciertos principios básicos que fueron establecidos primeramente por Newton. El establecimiento de estos principios involucra el concepto de derivada, y su aplicación requiere de la teoría de la integración. Sin analizar en detalle los principios de Newton, ilustraremos con algunos ejemplos sencillos cómo el cálculo se aplica en Mecánica.

a. Ley de movimiento de Newton

Nos restringiremos a la consideración de una única partícula, esto es, un punto en el cual se supone concentrada una masa m. Supondremos además que el movimiento se efectúa en el plano x, y, en el cual la posición de la partícula en el tiempo t está especificada por sus coordenadas $x = x(t)$, $y = y(t)$, o, equivalentemente, por su "vector de posición" $\mathbf{R} = \mathbf{R}(t) = (x(t), y(t))$. Un punto sobre una cantidad indica derivación con respecto al tiempo t. La *velocidad* y la *aceleración* de la partícula quedan entonces representadas por los vectores

$$\dot{\mathbf{R}} = (\dot{x}, \dot{y}) \qquad \text{y} \quad \ddot{\mathbf{R}} = (\ddot{x}, \ddot{y}).$$

En Mecánica se relaciona el movimiento de un punto con el concepto de *fuerzas* de dirección y magnitud definida que actúan sobre este punto. También puede describirse una fuerza mediante un vector $\mathbf{F} = (\rho, \sigma)$. El efecto de varias fuerzas $\mathbf{F}_1$, $\mathbf{F}_2, \ldots$ que actúan sobre la misma partícula es el mismo que el de una única fuerza $\mathbf{F}$, la *fuerza resultante,* que es simplemente la suma vectorial $\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2 + \cdots$ de las fuerzas individuales.

La ley fundamental de Newton establece que: *la masa m multiplicada por la aceleración es igual a la fuerza que actúa sobre la partícula;* en símbolos,

$$m\ddot{\mathbf{R}} = \mathbf{F}. \tag{43}$$

Si escribimos esta ecuación vectorial, que expresa la ley fundamental, en términos de las componentes de esos vectores, obtenemos el par equivalente de ecuaciones

$$m\ddot{x} = \rho, \qquad m\ddot{y} = \sigma. \tag{44}$$

Puesto que aceleración y fuerza difieren solamente en el factor positivo *m, la dirección de la aceleración es la misma que la de la fuerza.* Si ninguna fuerza actúa, esto es, $\mathbf{F} = \mathbf{O}$, la aceleración se anula, la velocidad es constante y x, y se vuelven funciones lineales de t. Esta es la *primera ley de Newton:* una partícula sobre la que no actúan fuerzas se mueve con velocidad constante a lo largo de una línea recta.

La ley de Newton $m\ddot{\mathbf{R}} = \mathbf{F}$, en el primer ejemplo, no es más que la definición cuantitativa del concepto de fuerza. El lado izquierdo de esta

relación puede determinarse observando el movimiento, mediante lo cual se obtiene entonces la fuerza.

Sin embargo, la ley de Newton tiene un significado más profundo debido al hecho de que en muchos casos podemos determinar la fuerza actuante a partir de otras consideraciones físicas, sin conocer el movimiento correspondiente. Esta ley fundamental no es entonces una *definición de fuerza* sino una relación a partir de la cual podemos esperar la determinación del movimiento. Esta manera de considerar las cosas al usar la ley de Newton muestra su utilidad en los numerosos casos en los que consideraciones físicas permiten expresar la fuerza **F** o sus componentes ρ, σ de manera explícita como funciones de la posición y la velocidad de la partícula y del tiempo t. La ley de movimiento no es entonces una tautología sino que proporciona dos ecuaciones que expresan $m\ddot{x}, m\ddot{y}$ en términos de x, y, $\dot{x}$, $\dot{y}$ y t, las llamadas *ecuaciones de movimiento*. Son éstas *ecuaciones diferenciales*, esto es, relaciones entre funciones y sus derivadas. La solución de estas ecuaciones diferenciales, o sea, el encontrar todas los pares de funciones $x(t)$, $y(t)$ para las cuales las ecuaciones de movimiento son válidas, dan todos los movimientos posibles de una partícula bajo la acción de la fuerza prescrita.

b. Movimiento de cuerpos en caída

El ejemplo más sencillo de una fuerza conocida es el de la *gravedad* que actúa sobre una partícula cercana a la superficie de la Tierra. Se sabe por observación directa que (despreciando los efectos de la resistencia del aire) todo cuerpo que cae tiene una aceleración dirigida verticalmente hacia abajo y que tiene la misma magnitud g para todos los cuerpos. Medida en metros por segundo (m.s^{-2}), g tiene aproximadamente el valor 9.80.[1] Si escogemos un sistema de coordenadas x, y en el cual el eje y apunte verticalmente hacia arriba en tanto que el eje x es horizontal, la aceleración $\ddot{\mathbf{R}} = (\ddot{x}, \ddot{y})$ tiene las componentes

$$\ddot{x} = 0, \qquad \ddot{y} = -g.$$

Por la ley fundamental de Newton, el vector **F** que representa la fuerza de gravedad que actúa sobre una partícula de masa m debe entonces ser

$$\mathbf{F} = (0, -mg).$$

Este vector fuerza está de la misma manera dirigido verticalmente hacia abajo; su magnitud, el *peso* del cuerpo cerca de la superficie de la Tierra, es mg.

[1] El valor preciso de g, que además de la atracción gravitacional incluye los efectos de la rotación de la Tierra, depende de la posición sobre ésta.

Cuando se cancela el factor m, las ecuaciones de movimiento de una partícula bajo la acción de la gravedad toman la forma

$$\ddot{x} = 0, \qquad \ddot{y} = -g.$$

A partir de estas ecuaciones podemos obtener fácilmente una descripción del movimiento más general posible para un cuerpo que cae. Integrando con respecto a t se obtiene

$$\ddot{x} = a, \qquad \ddot{y} = -gt + b,$$

donde a y b son constantes. Una integración posterior muestra entonces que

$$x = at + c, \qquad y = -\tfrac{1}{2}gt^2 + bt + d,$$

donde c y d son también constantes. Así, la solución general de nuestras ecuaciones de movimiento depende de cuatro constantes no especificadas, a, b, c, d. Podemos relacionar inmediatamente los valores de estas constantes para un movimiento particular con las *condiciones iniciales* de este movimiento. Si al tiempo inicial $t = 0$ la partícula está en el punto (x_0, y_0), entonces, haciendo $t = 0$, se encuentra

$$c = x_0, \qquad d = y_0.$$

La velocidad $\dot{\mathbf{R}} = (\dot{x}, \dot{y}) = (a, -gt + b)$ se reduce a (a, b) para $t = 0$. Así (c, d) y (a, b) representan respectivamente la posición y la velocidad iniciales de la partícula. Cualquier elección de estas condiciones iniciales origina en forma unívoca un movimiento.

En el caso $a \neq 0$, esto es, el caso en que la velocidad inicial no es vertical, podemos eliminar t y obtener una representación no paramétrica para la órbita de la partícula. Despejando t de la primera ecuación y substituyendo el resultado en la segunda, se llega a

$$y = -\frac{g}{2a^2}(x - c)^2 + \frac{b}{a}(x - c) + d.$$

Por lo tanto, la trayectoria es una parábola. Para $a = 0$ tenemos $x = c =$ constante, y todo el movimiento se efectúa a lo largo de una recta vertical.

c. Movimiento de una partícula restringido a una curva dada

En la mayoría de los problemas de mecánica las fuerzas que actúan sobre una partícula dependen de la posición y de la velocidad de dicha partícula. En general, las ecuaciones de movimiento son demasiado complicadas para permitirnos determinar todos los movimientos posibles. Una simplificación grande se logra si puede considerarse como conocida la curva C descrita por la partícula, de manera que sólo tenga que determinarse el

movimiento a lo largo de la curva. En una variedad muy grande de problemas mecánicos la partícula está *restringida* (mediante algún artificio mecánico) a moverse sobre una curva dada C. El ejemplo más simple lo constituye el péndulo plano, en el que una masa m está sujeta a un punto P_0 por medio de un hilo inextensible de longitud L y se mueve bajo la acción de la gravedad sobre una circunferencia de radio L con centro en P_0.

Usando el arco de longitud s como parámetro a lo largo de la curva C, ésta queda determinada por las ecuaciones $x = x(s)$, $y = y(s)$. Encontrar el movimiento de la partícula a lo largo de C equivale entonces a encontrar s como función de t. Una ecuación de movimiento válida a lo largo de la curva se encuentra de la manera siguiente.

Multiplíquense escalarmente ambos lados de la fórmula de Newton $m\ddot{\mathbf{R}} = \mathbf{F}$ por un vector ξ:

$$m\ddot{\mathbf{R}} \cdot \xi = \mathbf{F} \cdot \xi.$$

Si se escoge ξ igual al vector de longitud 1 cuya dirección es la de la tangente a C en el sentido de las s crecientes, esto es, $\xi = d\mathbf{R}/ds$, la expresión $\mathbf{F} \cdot \xi = f$ dará la ***componente tangencial*** de la fuerza, o ***fuerza que actúa en la dirección del movimiento.*** De acuerdo con la ecuación (41) de la p. 414, la componente tangencial $\ddot{\mathbf{R}} \cdot \xi$ de la aceleración es justamente $dv/dt = d^2s/dt^2$, esto es, la aceleración de la partícula a lo largo de la curva. La ley de Newton conduce entonces a la fórmula

$$m\ddot{s} = f, \tag{45}$$

o sea, ***la masa de la partícula multiplicada por su aceleración a lo largo de la trayectoria es igual a la fuerza que actúa en la dirección del movimiento.***

Al aplicar esta ecuación a una partícula *restringida* a moverse sobre una curva C, suponemos que las restricciones no contribuyen a la fuerza f.[1] Para una fuerza $\mathbf{F} = (\rho, \sigma)$ la ecuación (44) de la p. 415 conduce a

$$f = \rho\frac{dx}{ds} + \sigma\frac{dy}{ds}; \tag{46}$$

puesto que el vector ξ tiene como componentes $\dfrac{dx}{ds}, \dfrac{dy}{ds}$ (ver p. 412). Para una curva C conocida los cosenos directores $\dfrac{dx}{ds}$ y $\dfrac{dy}{ds}$ de la tangente pueden considerarse como funciones conocidas de s. De la misma manera, si la fuerza $\mathbf{F} = 2(\rho, \sigma)$ depende solamente de la posición de la partícula, f será una función conocida de s. El movimiento de la partícula a lo largo de C

[1] Realmente, el mecanismo de restricción tiene que proporcionar una fuerza que mantenga a la partícula sobre C (en el péndulo simple la proporciona la tensión del hilo). Suponemos que esta fuerza de "reacción" es perpendicular a la curva y así no tiene componente tangencial; y éste es el caso cuando la partícula se desliza sin fricción sobre la curva.

tiene entonces que ser determinado a partir de la ecuación diferencial relativamente sencilla $m\ddot{s} = f(s)$.

Específicamente, para la fuerza gravitacional $\mathbf{F} = (0, -mg)$ se tiene

$$f = -mg\frac{dy}{ds}. \tag{46a}$$

Así, la ecuación de movimiento de una partícula restringida a moverse bajo la acción de la gravedad sobre una curva C se transforma en

$$\frac{d^2s}{dt^2} = -g\frac{dy}{ds}. \tag{47}$$

Si α denota el ángulo de inclinación de la curva, tenemos $dy/ds = \operatorname{sen}\alpha$ (ver Fig. 4.43), y la ecuación de movimiento se convierte en

$$\frac{d^2s}{dt^2} = -g \operatorname{sen}\alpha$$

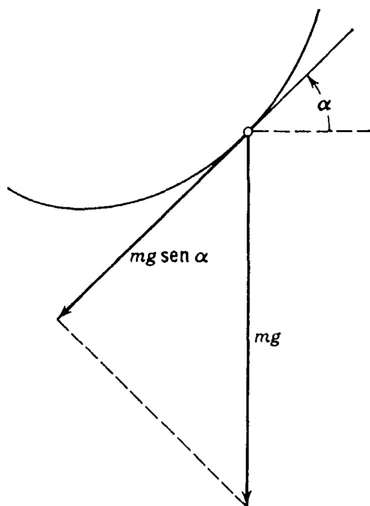


Figura 4.43 Movimiento sobre una curva dada, bajo la acción de la gravedad.

Para una partícula restringida a moverse sobre una circunferencia de radio L con centro en el origen ("péndulo simple"), es

$$x = L \operatorname{sen}\theta, \quad y = -L\cos\theta,$$

donde $\theta = s/L$ es el ángulo polar medido a partir de la vertical que apunta hacia abajo. En este caso (ver Fig. 4.44) es $\alpha = \theta$, por lo que

$$\frac{d^2s}{dt^2} = -g\frac{dy}{d\theta}\frac{d\theta}{ds} = -g\operatorname{sen}\theta,$$

o bien

$$\frac{d^2\theta}{dt^2} = -\frac{g}{L}\operatorname{sen}\theta.$$

4.5 Caída libre de un cuerpo venciendo la resistencia del aire

Comenzaremos con dos ejemplos relativos al movimiento de una partícula a lo largo de una recta. Consideraremos solamente los casos en que la fuerza actúa en la dirección de la recta, de manera que no es necesario un mecanismo de restricción.

La trayectoria de un cuerpo que cae libremente hacia abajo puede describirse mediante las expresiones $x =$ constante, $y = s$. Si la única fuerza presente es la de gravedad, se tiene la ecuación de movimiento

$$m\ddot{s} = -mg.$$

Para una partícula que se suelta en el tiempo $t = 0$ desde la altura $y_0 = s_0$ con velocidad inicial v_0 (considerada positiva si es hacia arriba), se encuentra por integración que

$$s = -\tfrac{1}{2}gt^2 + v_0t + s_0.$$

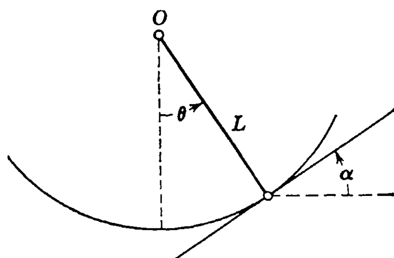


Figura 4.44 El péndulo simple.

Si se desea tomar en cuenta el efecto de la *fricción* o *resistencia del aire* que actúa sobre la partícula, se la tendrá que considerar como una fuerza de sentido opuesto al del movimiento y respecto a la cual pueden hacerse ciertas consideraciones de tipo físico.[1] Se obtendrán resultados para diferentes hipótesis físicas: *a)* la resistencia es proporcional a la velocidad y está dada por una expresión de la forma $-r\dot{s}$, donde r es una constante positiva; *b)* la resistencia es proporcional al cuadrado de la velocidad y es de la forma $-r\dot{s}^2$ para $\dot{s}$ positiva, y de la forma $r\dot{s}^2$ para $\dot{s}$ negativa. De acuerdo con la ley de Newton se obtienen las ecuaciones de movimiento

$$(a) \qquad m\ddot{s} = -mg - r\dot{s},$$

$$(b) \qquad m\ddot{s} = -mg + r\dot{s}^2,$$

[1] Estas hipótesis deben escogerse adecuadas al sistema físico particular que se esté considerando; por ejemplo, la ley de resistencia para bajas velocidades no es la misma que para velocidades altas (tales como velocidades de bala).

donde se ha supuesto para (b) que el cuerpo está cayendo $(\dot{s} < 0)$. Si se considera, en primer lugar, $\dot{s} = v(t)$ como la función buscada, se tiene

$$(a) \qquad m\dot{v} = -mg - rv,$$

$$(b) \qquad m\dot{v} = -mg + rv^2.$$

En vez de usar estas ecuaciones para determinar v como función de t, se determina t como función de v escribiendo dichas ecuaciones diferenciales en la forma

$$(a) \qquad \frac{dt}{dv} = -\frac{1}{g(1 + k^2 v)},$$

$$(b) \qquad \frac{dt}{dv} = -\frac{1}{g(1 - k^2 v^2)}$$

donde se ha hecho $\sqrt{r/mg} = k$. Con la ayuda de los métodos dados en el capítulo 3 pueden realizarse inmediatamente las integraciones, obteniéndose

$$(a) \qquad t = -\frac{1}{gk^2} \log (1 + k^2 v) + t_0,$$

$$(b) \qquad t = \frac{1}{2gk} \log \frac{1 - kv}{1 + kv} + t_0.$$

Resolviendo para v estas ecuaciones se tiene

$$(a) \qquad v = -\frac{1}{k^2} (1 - e^{-gk^2(t-t_0)}).$$

$$(b) \qquad v = -\frac{1}{k} \frac{1 - e^{-2gk(t-t_0)}}{1 + e^{-2gk(t-t_0)}} = -\frac{1}{k} \tanh [gk(t - t_0)]$$

Estas ecuaciones revelan ya una propiedad importante del movimiento. La velocidad no crece con el tiempo más allá de toda cota, sino que tiende a un límite definido que depende de la masa m y de la constante r (la que, a su vez, depende de la forma del cuerpo que cae y de la densidad del aire). En efecto,

$$(a) \qquad \lim_{t\to\infty} v(t) = -\frac{1}{k^2} = -\frac{mg}{r},$$

$$(b) \qquad \lim_{t\to\infty} v(t) = -\frac{1}{k} = -\sqrt{\frac{mg}{r}}$$

La resistencia de fricción para la velocidad límite contrarresta justamente al atracción gravitacional. Una segunda integración realizada en las expresiones para $v(t) = \dot{s}$, nuevamente con la ayuda de los métodos del capítulo 3, conduce a los resultados (que pueden verificarse por derivación)

$$(a) \qquad s(t) = -\frac{1}{k^2} (t - t_0) - \frac{1}{gk^4} (e^{-gk^2(t-t_0)} - 1) + c$$

$$(b) \qquad s(t) = -\frac{1}{gk^2}\log[\cosh kg(t - t_0)] + c,$$

donde c es una constante de integración. Aquí t_0 es el instante para el cual la velocidad de la partícula tendría que ser 0, y c es la altura correspondiente. Las dos constantes c y t_0 pueden también ser fácilmente relacionadas con la velocidad y la posición en cualquier otro instante t_1 si se consideran aquellas cantidades como condiciones iniciales.

4.6 El tipo más simple de vibración elástica: movimiento de un resorte

Como un segundo ejemplo —de mayor significación— considérese el caso de una partícula que se mueve sobre el eje x bajo la acción de una fuerza elástica que la atrae hacia el origen. Por lo que respecta a la fuerza elástica se considerará que su magnitud es proporcional a la distancia desde el origen y que está siempre dirigida hacia ese punto. En otras palabras, se tomará la fuerza igual a $-kx$, donde el coeficiente k es una medida de la rigidez de la conexión elástica. Puesto que k es positiva, la fuerza es negativa cuando x es positiva y positiva cuando x es negativa. La ley de Newton en este caso establece que

$$(48) \qquad m\ddot{x} = -kx.$$

Esta ecuación diferencial por sí sola no determina completamente el movimiento, sino que para un instante dado, por ejemplo $t = 0$, pueden fijarse arbitrariamente la posición $x(0) = x_0$ y la velocidad $\dot{x}(0) = v_0$ iniciales; esto es, en lenguaje físico, que la partícula puede iniciar su movimiento desde cualquier posición y con cualquier velocidad, quedando determinado el movimiento subsecuente por la ecuación diferencial. Matemáticamente esto queda expresado por el hecho de que la solución general de la ecuación diferencial contiene dos constantes de integración cuyos valores, al principio indeterminados, pueden encontrarse mediante las condiciones iniciales. Esto se demostrará de inmediato.

Podemos establecer fácilmente dicha solución de una manera directa. Si hacemos $\omega = \sqrt{k/m}$, nuestra ecuación diferencial se convierte en $d^2x/dt^2 = -\omega^2 x$. La substitución $\tau = \omega t$ de la variable independiente reduce esta ecuación a la forma $d^2x/d\tau^2 = -x$, discutida en el capítulo 3, p. 331. Así la ecuación diferencial es satisfecha por todas las funciones

$$x(t) = c_1 \cos \omega t + c_2 \operatorname{sen} \omega t,$$

lo cual puede también verificarse de inmediato por derivación (c_1 y c_2 denotan constantes arbitriamente escogidas). En el capítulo 3, p. 332, vimos que no existen otras soluciones para dicha ecuación diferencial, y, por lo tanto, cualquier movimiento originado por la acción de la fuerza

elástica está dado por esa expresión. Esta puede fácilmente ponerse en la forma

$$x(t) = a \operatorname{sen} \omega(t - \delta) = -a \operatorname{sen} \omega \delta \cos \omega t + a \cos \omega \delta \operatorname{sen} \omega t.$$

Sólo se necesita escribir $-a \operatorname{sen} \omega \delta = c_1$ y $a \cos \omega \delta = c_2$, introduciendo así las nuevas constantes a y δ en vez de c_1 y c_2. Movimientos de este tipo son llamados *sinusoidales* o *armónicos simples*. Estos movimientos son periódicos: cualquier *estado* [esto es, posición $x(t)$ y velocidad $\dot{x}(t)$] se repite después de un tiempo $T = 2\pi/\omega$, llamado *período,* pues las funciones $\operatorname{sen} \omega t$ y $\cos \omega t$ tienen período T. El número a se llama el *desplazamiento máximo* o *amplitud* de la oscilación. El número $1/T = \omega/2\pi$ se llama la *frecuencia* de la oscilación; mide el número de oscilaciones por unidad de tiempo. En el capítulo 8 regresaremos a la teoría de las oscilaciones.

4.7 Movimiento sobre una curva dada

a. La ecuación diferencial y su solución

Volvamos ahora a la forma general del problema relativo al movimiento a lo largo de una cierta curva bajo la acción de una fuerza $mf(s)$ arbitraria pero prefijada. Determinaremos $s(t)$ como función de t usando la ecuación diferencial [Ec. (45), p. 419]

$$\ddot{s} = f(s),$$

donde $f(s)$ es una función dada.[1] Esta ecuación diferencial en s puede resolverse completamente mediante el siguiente artificio.

Consideremos cualquier función primitiva $F(s)$ de $f(s)$, esto es $F'(s) = f(s)$, y multipliquemos ambos lados de la ecuación $\ddot{s} = f(s) = F'(s)$ por $\dot{s}$. Podemos entonces escribir el miembro de la izquierda en la forma $d(\dot{s}^2/2)\,dt$, como se ve de inmediato derivando la expresión $\dot{s}^2$. Por otra parte, si en $F(s)$ consideramos la cantidad s como una función de t, por la regla de la cadena de la derivación, el lado de la derecha, $F'(s)\dot{s}$, es la derivada de $F(s)$ con respecto al tiempo t. Por lo tanto, se tiene inmediatamente

$$\frac{d}{dt}\left(\frac{1}{2}\dot{s}^2\right) = \frac{d}{dt}F(s),$$

o bien, integrando,

$$\tfrac{1}{2}\dot{s}^2 = F(s) + c,$$

donde c denota una constante por determinar.

Hemos, en este punto, llegado a una ecuación que sólo involucra la función $s(t)$ y su *primera* derivada. (Más adelante interpretaremos esta

[1] Nuestra ecuación original de movimiento sobre una curva era $m\ddot{s} = f(s)$; sin embargo, podemos siempre escribir la función $f(s)$ en la forma $mf(s)$, obteniendo la forma más simple usada aquí para la ecuación.

ecuación como la que expresa la conservación de la energía durante el movimiento.) Escríbase esta ecuación en la forma $ds/dt = \sqrt{2[F(s) + c]}$. Vemos que no puede encontrarse inmediatamente por integración s como función de t. Sin embargo, obtenemos una solución del problema si primero nos concretamos a encontrar la función inversa $t(s)$, esto es, el tiempo que tarda la partícula en alcanzar una posición definida s. Para $t(s)$ tenemos la ecuación

$$\frac{dt}{ds} = \frac{1}{\sqrt{2[F(s) + c]}};$$

y así, la derivada de la función $t(s)$ es conocida y se tiene

$$t = \int \frac{ds}{\sqrt{2[F(s) + c]}} + c_1,$$

donde c_1 es otra constante de integración. El problema se habrá resuelto tan pronto como se realice esta última integración, pues aunque no se determine la posición s como una función de t, se encuentra, por el contrario, el tiempo t como una función de la posición s. El hecho de que todavía se disponga de dos constantes de integración, c y c_1, nos permite hacer que la solución general satisfaga condiciones iniciales específicas.

La discusión general puede ilustrarse con nuestro primer ejemplo sobre vibraciones elásticas si identificamos x con s. En este caso, $f(s) = -\omega^2 s$ y, por ejemplo, $F(s) -\frac{1}{2}\omega^2 s^2$. Obtenemos entonces

$$\frac{dt}{ds} = \frac{1}{\sqrt{2c - \omega^2 s^2}},$$

y además,

$$t = \int \frac{ds}{\sqrt{2c - \omega^2 s^2}} + c_1.$$

Esta integral puede evaluarse fácilmente introduciendo $\omega s/\sqrt{2c}$ como una nueva variable: así obtenemos

$$t = \frac{1}{\omega} \operatorname{arc\,sen} \frac{\omega s}{\sqrt{2c}} + c_1,$$

o bien, tomando la función inversa,

$$s = \frac{\sqrt{2c}}{\omega} \operatorname{sen} \omega(t - c_1).$$

Llegamos así a exactamente la misma fórmula obtenida anteriormente para la solución.

Este ejemplo ilustra también lo que significan las constantes de integración y cómo se determinan. Si, por ejemplo, se requiere que en el

tiempo $t = 0$ la partícula esté en el punto $s = 0$ y que su velocidad en ese instante sea $\dot{s}(0) = 1$, se obtienen las dos ecuaciones

$$0 = \frac{\sqrt{2c}}{\omega} \operatorname{sen} \omega c_1, \qquad 1 = \sqrt{2c} \cos \omega c_1,$$

a partir de las cuales se encuentra que las constantes tienen los valores $c_1 = 0$, $c = \frac{1}{2}$. Las constantes de integración c y c_1 pueden determinarse exactamente de la misma manera cuando se prescriben arbitrariamente la posición s_0 y la velocidad $\dot{s}_0$ iniciales (en el tiempo $t = 0$).

b. Partícula que se desliza hacia abajo sobre una curva

El caso de una partícula que se desliza sin fricción sobre una curva bajo la acción de la gravedad puede tratarse en forma muy simple mediante el método descrito anteriormente. Encontramos ya en la p. 419 la ecuación de movimiento correspondiente a este caso:

$$\ddot{s} = -g\frac{dy}{ds},$$

donde los puntos indican derivación con respecto a t. El miembro derecho de esta ecuación es una función conocida de s, puesto que conocemos la curva y podemos por consiguiente considerar x e y como funciones conocidas de s.

Multipliquemos ambos miembros de esta ecuación por $\dot{s}$, como en la última sección. El miembro izquierdo se convierte entonces en la derivada con respecto a t de $\frac{1}{2}\dot{s}^2$. Si en la expresión $y(s)$ consideramos s como una función de t, el miembro derecho de nuestra ecuación es la derivada con respecto a t de $-gy$. Integrando tendremos entonces

$$\tfrac{1}{2}\dot{s}^2 = -gy + c,$$

donde c es una constante de integración. Para encontrar el significado de esta constante supongamos que en el tiempo $t = 0$ nuestra partícula está en el punto de la curva con coordenadas x_0 e y_0, y que en ese instante su velocidad es cero, esto es, $\dot{s}(0) = 0$. Entonces, haciendo $t = 0$ tenemos inmediatamente $-gy_0 + c = 0$, de manera que

$$\tfrac{1}{2}\dot{s}^2 = g(y_0 - y).$$

Puesto que $\dot{s}^2$ nunca es negativo, la altura y de la partícula jamás excede al valor y_0 y sólo es igual a él cuando la velocidad es cero. Mientras más abajo se encuentre la partícula, mayor será su velocidad. Ahora, en vez de considerar s como una función de t, consideraremos la función inversa $t(s)$. Para esta función obtenemos inmediatamente

$$\frac{dt}{ds} = \pm \frac{1}{\sqrt{2g(y_0 - y)}}$$

que es equivalente a

$$t = c_1 \pm \int \frac{ds}{\sqrt{2g(y_0 - y)}},$$

donde c_1 es una nueva constante de integración. En relación con el signo de la raíz cuadrada, que es el mismo que el de $\dot{s}$, observamos que si la partícula se mueve a lo largo de un arco que en todas partes está por debajo de y_0 excepto en los extremos, el signo no puede cambiar. Pues el signo de $\dot{s}$ sólo puede cambiar donde $\dot{s} = 0$, esto es, donde $y - y_0 = 0$. De esta manera, la partícula solamente puede "regresar" en puntos de la curva de máxima elevación. En vez de usar la longitud de arco s como parámetro de referencia para la curva, puede usarse cualquier otro parámetro θ tal que $x = \phi(\theta)$, $y = \psi(\theta)$. Introduciendo θ como variable independiente se obtiene

$$t = c_1 \pm \int \frac{ds}{d\theta} \frac{d\theta}{\sqrt{2g(y_0 - y)}} = c_1 \pm \int \frac{x'^2 + y'^2}{2g(y_0 - y)} d\theta,$$

donde las funciones $x' = \phi'(\theta)$, $y' = \psi'(\theta)$, e $y = \psi(\theta)$ son conocidas. Para determinar la constante de integración c_1 observamos que para $t = 0$ el parámetro θ tendrá un valor θ_0. Esto nos da inmediatamente la solución en la forma

$$(49) \qquad t = \pm \int_{\theta_0}^{\theta} \sqrt{\frac{x'^2 + y'^2}{2g(y_0 - y)}}\, d\theta.$$

Nótese que esta función representa el tiempo que tarda la partícula para moverse del punto con valor paramétrico θ_0 al punto con valor paramétrico θ. La función inversa $\theta(t)$ correspondiente nos permite describir com-

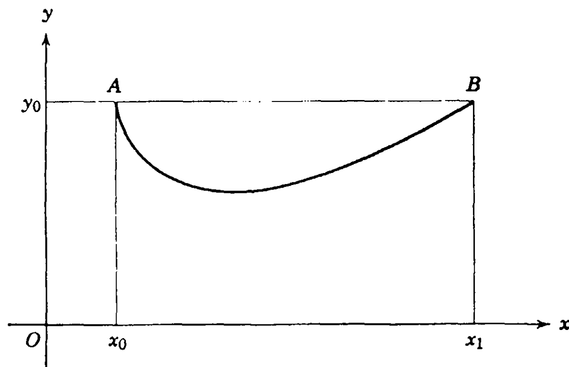


Figura 4.45

pletamente el movimiento: para cada instante t podemos determinar el punto $x = \phi[\theta(t)]$, $y = \psi[\theta(t)]$ por el que la partícula está pasando.

c. Discusión del movimiento

De las ecuaciones recién encontradas puede deducirse, aún sin una expresión explícita para el resultado de la integración, la naturaleza general del movimiento mediante un simple razonamiento intuitivo. Supongamos que nuestra curva es del tipo mostrado en la Fig. 4.45, esto es, que consiste de un arco convexo hacia abajo; y tomemos s creciente de izquierda a derecha. Si inicialmente soltamos la partícula desde el punto A de coordenadas $x_0 = \phi(\theta_0)$, $y_0 = \psi(\theta_0)$, correspondientes a $\theta = \theta_0$, la velocidad aumenta, pues la aceleración $\ddot{s}$ es positiva. La partícula viaja de A al punto más bajo con velocidad siempre creciente. Sin embargo, después de pasar por el punto más bajo la aceleración es negativa ya que el miembro derecho, $-g\, dy/ds$, de la ecuación de movimiento es negativo. Por lo tanto, la velocidad decrece. De la ecuación $\dot{s}^2 = 2g(y_0 - y)$ se ve de inmediato que la velocidad alcanza el valor cero cuando la partícula llega al punto B cuya altura es la misma que la de la posición inicial A. Puesto que la aceleración es todavía negativa, el movimiento debe invertirse en ese punto y la partícula oscilará de regreso hacia A; y esta acción se repetirá indefinidamente. (El lector recordará que la fricción no se ha tomado en cuenta.) En este movimiento oscilatorio el tiempo que tarda el punto en regresar de B a A debe claramente ser el mismo que el que tarda en moverse de A a B, pues para alturas iguales se tienen valores iguales de $|\dot{s}|$. Si denotamos con T el tiempo requerido para un viaje completo de A a B y viceversa, el movimiento obviamente será periódico con período T. Si θ_0 y θ_1 son los valores del parámetro correspondientes a los puntos A y B, respectivamente, el semiperíodo está dado por la expresión

$$\text{(50)} \qquad \begin{aligned} \frac{T}{2} &= \frac{1}{\sqrt{2g}} \left| \int_{\theta_0}^{\theta_1} \frac{x'^2 + y'^2}{y_0 - y}\, d\theta \right| \\ &= \frac{1}{\sqrt{2g}} \left| \int_{\theta_0}^{\theta_1} \frac{\phi'^2(\theta) + \psi'^2(\theta)}{\psi(\theta_0) - \psi(\theta)}\, d\theta \right|. \end{aligned}$$

Si θ_2 es el valor del parámetro correspondiente al punto más bajo de la curva, el tiempo que la partícula tarda en caer de A hasta ese punto es

$$\frac{1}{\sqrt{2g}} \left| \int_{\theta_0}^{\theta_2} \sqrt{\frac{x'^2 + y'^2}{y_0 - y}}\, d\theta \right|.$$

d. El péndulo ordinario

El ejemplo más sencillo está dado por el llamado péndulo simple. En este caso la curva a considerar es una circunferencia de radio fijo L:

$$x = L \operatorname{sen} \theta, \qquad y = -L \cos \theta,$$

donde el ángulo θ se mide en el sentido positivo a partir de la posición de reposo. De la expresión general (50), y usando el teorema de adición para el coseno, se obtiene inmediatamente

$$T = \sqrt{\frac{2L}{g}} \int_{-\theta_0}^{\theta_0} \frac{d\theta}{\sqrt{\cos \theta - \cos \theta_0}} = \sqrt{\frac{L}{g}} \int_{-\theta_0}^{\theta_0} \frac{d\theta}{\sqrt{\operatorname{sen}^2 \frac{\theta_0}{2} - \operatorname{sen}^2 \frac{\theta}{2}}},$$

donde θ_0 $(0 < \theta_0 < \pi)$ denota la amplitud de oscilación del péndulo, esto es, la posición angular desde la que se suelta la partícula con velocidad cero en el tiempo $t = 0$.[1] Mediante la substitución

$$u = \frac{\operatorname{sen} (\theta/2)}{\operatorname{sen} (\theta_0/2)}, \qquad \frac{du}{d\theta} = \frac{\cos (\theta/2)}{2 \operatorname{sen} (\theta_0/2)},$$

nuestra expresión para el período de oscilación del péndulo se transforma en

$$T = 2 \sqrt{\frac{L}{g}} \int_{-1}^{1} \frac{du}{\sqrt{(1 - u^2)\left(1 - u^2 \operatorname{sen}^2\left(\frac{\theta_0}{2}\right)\right)}}.$$

Tenemos así expresado el período de oscilación del péndulo mediante una *integral elíptica* (ver p. 318).

Si se supone que la amplitud de la oscilación es pequeña, de manera que con suficiente aproximación se puede reemplazar el segundo factor del subradical por 1, se obtiene la expresión

$$2 \sqrt{\frac{L}{g}} \int_{-1}^{1} \frac{du}{\sqrt{1 - u^2}}$$

como una aproximación para el período de oscilación. Esta última integral puede evaluarse usando la fórmula 13 de nuestra tabla de integrales (p. 284); así se obtiene la expresión $2\pi \sqrt{L/g}$ como un valor aproximado para T. Hasta este orden de aproximación, el período es independiente de θ_0, o sea, de la amplitud de oscilación del péndulo. Es claro que el

[1] Hemos supuesto aquí que la velocidad se hace cero en algún momento durante el movimiento. Esto excluye el tipo de movimiento de salto del péndulo, para el cual θ no es periódica y varía monótonamente para todo t.

período exacto es mayor y aumenta con θ_0. Puesto que en el intervalo de integración

$$1 \geq 1 - u^2 \operatorname{sen}^2 \frac{\theta_0}{2} \geq 1 - \operatorname{sen}^2 \frac{\theta_0}{2} = \cos^2 \frac{\theta_0}{2},$$

se encuentran las siguientes estimaciones para el período

$$2\pi \sqrt{\frac{L}{g}} \leq T \leq \frac{1}{\cos(\theta_0/2)} 2\pi \sqrt{\frac{L}{g}}.$$

Para ángulos $\theta_0 < 10°$ se tiene $1/(\cos \theta_0/2) \leq \operatorname{seg} 5° < 1.004$, de manera que el período estará dado por la fórmula $2\pi\sqrt{L/g}$ con un error relativo menor que el $\frac{1}{2}\%$. Para una aproximación más fina de la integral elíptica que expresa T véase la sección 7.6*f*.

e. El péndulo cicloidal

El hecho de que el período de oscilación de un péndulo ordinario no es estrictamente independiente de la amplitud de oscilación hizo que Christian Huygens, en sus constantes esfuerzos por construir relojes precisos, buscase una curva C para la cual el período de oscilación fuese independiente de la posición sobre C desde la cual la partícula oscilante inicia su movimiento.[1] Huygens reconoció que la cicloide es tal curva.

Para que una partícula sea realmente capaz de oscilar sobre una cicloide, las cúspides de esta curva deben apuntar en dirección opuesta a la de la gravedad; esto es, la ciclode considerada anteriormente (p. 347) debe ser reflejada con respecto del eje x (cfr. Fig. 4.2, p. 348). Por consiguiente, escribimos las ecuaciones de la cicloide en la forma

$$x = a(\theta + \pi + \operatorname{sen} \theta),$$

$$y = -a(1 + \cos \theta),$$

la cual involucra también un cambio del parámetro t por $\theta + \pi$ (Fig. 4.46). El tiempo que la partícula tarda en ir de un punto con altura

$$y_0 = -a(1 + \cos \theta_0) \qquad (0 < \theta_0 < \pi)$$

al punto más bajo y luego subir nuevamente a la altura y_0 es, por la fórmula (50) de la p. 428,

$$\frac{T}{2} = \sqrt{\frac{1}{2g}} \int_{-\theta_0}^{\theta_0} \sqrt{\frac{x'^2 + y'^2}{y_0 - y}}\, d\theta = \sqrt{\frac{2a}{g}} \int_{-\theta_0}^{\theta_0} \frac{\cos(\theta/2)}{\sqrt{\cos\theta - \cos\theta_0}}\, d\theta.$$

---

[1] Las oscilaciones se dice entonces que son *isócronas*.

Usando exactamente las mismas substituciones que para el período del péndulo simple se llega a la integral

$$\frac{T}{2} = 2\sqrt{\frac{a}{g}}\int_{-1}^{1}\frac{du}{\sqrt{1-u^2}},$$

y, por consiguiente,

$$T = 4\pi\sqrt{\frac{a}{g}}.$$

El período de oscilación T, por lo tanto, es, sin duda, independiente de la amplitud θ_0. En la p. 446 se describirá una manera sencilla de lograr, mediante un hilo, que una partícula se mueva realmente sobre una cicloide.

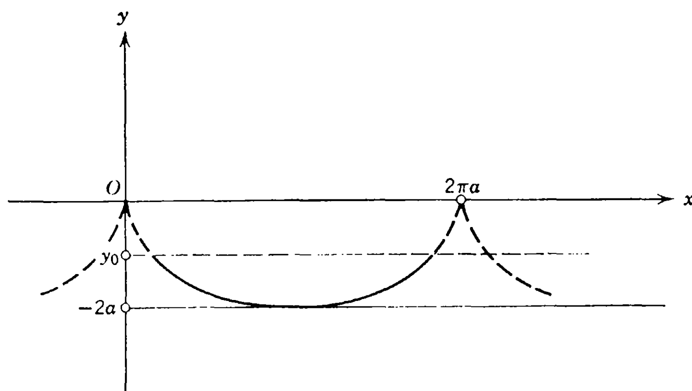


Figura 4.46 Trayectoria descrita por un péndulo cicloidal.

*4.8 Movimiento de un campo gravitacional

Como un ejemplo de movimiento no restringido, consideremos una partícula que se mueve en el campo gravitacional de una masa atractiva.

a. Ley universal de la gravitación de Newton

La descripción del movimiento de los planetas dada por Kepler, basada en las precisas observaciones de Tycho Brahe, condujo a Newton a formular su ley general de la atracción gravitacional entre dos partículas cualesquiera. Sean $P_0 = (x_0, y_0)$ y $P = (x, y)$ dos partículas de masas m_0 y m, respectivamente. Sea $r = \sqrt{(x - x_0)^2 + (y - y_0)^2}$ la distancia entre las partículas. Entonces P_0 ejerce sobre P una fuerza $\mathbf{F}$ que tiene la dirección de $\overrightarrow{PP_0}$ y la magnitud $|\mathbf{F}| = \gamma m_0 m/r^2$, donde γ es la "*constante universal de la gravitación.* Puesto que entonces $\mathbf{F}$ solamente puede diferir en un factor positivo del vector $\overrightarrow{PP_0}$, cuya magnitud es r, debe tenerse

$$\mathbf{F} = \frac{\gamma m_0 m}{r^3}\overrightarrow{PP_0} = \left(\frac{\gamma m_0 m(x_0 - x)}{r^3} \quad \frac{\gamma m_0 m(y_0 - y)}{r^3}\right).$$

Esta ley de atracción se refiere a *partículas,* esto es, a cuerpos que pueden considerarse concentrados en puntos, despreciando la extensión real de los mismos (Fig. 4.47). La validez de tal hipótesis es lo suficientemente plausible para cuerpos celestes, cuyas distancias mutuas son enormes comparadas con sus diámetros. Newton extendió considerablemente el rango de aplicación de esta ley demostrando que la misma también describe la

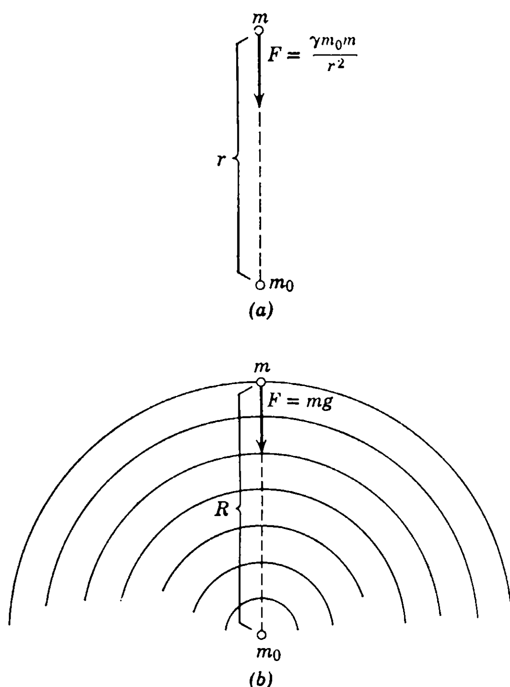


Figura 4.47 *a*) Atracción newtoniana entre dos partículas. *b*) Atracción gravitacional de la Tierra.

atracción que un cuerpo de masa m_0 y extensión no nula ejerce sobre una partícula de masa *m,* suponiendo que el cuerpo es una esfera cuya densidad es constante o, más generalmente, suponiendo que el cuerpo está constituido por capas esféricas concéntricas de densidad constante. En este caso la atracción del cuerpo sobre una partícula *P* situada fuera de él es la misma que si toda la masa m_0 del cuerpo estuviese concentrada en su centro P_0 (Fig. 4.47). La Tierra puede, con bastante aproximación, considerarse como constituida por capas concéntricas de densidad constante, de manera que su atracción sobre una partícula de masa *m* situada en la super-

ficie terrestre está diriigda hacia el centro P_0 de la Tierra (esto es, verticalmente hacia abajo para un observador) y tiene una magnitud igual a $\gamma m_0 m/R^2$, donde R es el radio de la Tierra y m_0 su masa. Se puede identificar $\gamma m_0 m/R^2$ con mg, siendo g la aceleración gravitacional (ver p. 416). En otras palabras, $g = \gamma m_0/R^2$.

A partir de la ley fundamental de Newton se encuentran las ecuaciones de movimiento para una partícula P de masa m que se mueve bajo la influencia de la atracción de una masa m_0 situada en P_0:

$$x = \frac{\gamma m_0(x_0 - x)}{r^3}, \qquad y = \frac{\gamma m_0(y_0 - y)}{r^3}.$$

Supóngase ahora, para simplificar, que m_0 es mucho mayor que m, que los efectos de la atracción de P sobre P_0 pueden despreciarse y que P_0 puede considerarse en reposo. Esta podría ser, por ejemplo, la situación de una pareja de cuerpos como el Sol y un planeta, o bien la Tierra y un objeto sobre su superficie. Tomando el origen de coordenadas en P_0 se tienen para $P = (x, y)$ las ecuaciones de movimiento

$$(51) \qquad \ddot{x} = -\frac{\gamma m_0 x}{r^3}, \qquad \ddot{y} = -\frac{\gamma m_0 y}{r^3}$$

con $r = \sqrt{x^2 + y^2}$.

b. Movimiento circular en torno al centro de atracción

No se intentará encontrar la solución más general de estas ecuaciones diferenciales (lo que correspondería, como es bien sabido, al movimiento sobre una trayectoria en forma de sección cónica con un foco en el centro de atracción). En vez de ello, se considerarán solamente los tipos más simples de movimiento compatibles con estas ecuaciones, como son los movimientos circulares uniformes en torno al origen y los movimientos a lo largo de un radio que parte de ese mismo punto. Para un movimiento circular uniforme de P sobre la circunferencia de radio a con centro en el origen, se tiene $r = a$ y

$$x = a \cos \omega t, \qquad y = a \operatorname{sen} \omega t,$$

donde ω es una constante. El *período* T del movimiento, esto es, el tiempo que tarda P en regresar a la misma posición, es $T = 2\pi/\omega$. Para las componentes de la velocidad se encuentra

$$\dot{x} = -a\omega \operatorname{sen} \omega t, \qquad \dot{y} = a\omega \cos \omega t,$$

de manera que la velocidad de P sobre su órbita es

$$(52) \qquad v = \sqrt{\dot{x}^2 + \dot{y}^2} = a\omega = \frac{2\pi a}{T}.$$

La aceleración de P tiene por componentes

$$\ddot{x} = -a\omega^2 \cos \omega t = -\omega^2 x, \qquad \ddot{y} = -a\omega^2 \operatorname{sen} \omega t = -\omega^2 y.$$

Es claro entonces que las ecuaciones de movimiento (51) son satisfechas si

$$\omega^2 = \frac{\gamma m_0}{a^3},$$

o bien

(53) $$a^3 = \frac{\gamma m_0}{\omega^2} = \frac{\gamma m_0}{4\pi^2} T^2.$$

Esta es precisamente la *tercera ley de Kepler* para el caso especial de movimiento circular, según la cual *los cubos de las distancias de los planetas al Sol son proporcionales a los cuadrados de sus períodos.*

Pueden darse algunas ilustraciones sencillas de la ley de Kepler para el caso en que el cuerpo atractivo es la Tierra, con m_0 su masa y R su radio. Observando que en este caso $\gamma m_0 = gR^2$, se tiene

$$a^3 = \frac{gR^2}{4\pi^2} T^2.$$

Para un satélite que circunvale la Tierra al nivel de la cima de un árbol (despreciando, por supuesto, la resistencia del aire) se tiene $a = R \sim 3963$ millas. A partir de nuestra fórmula se encuentra para el período del satélite el valor

$$T = 2\pi \sqrt{\frac{R}{g}} \sim 1.4 \text{ horas}$$

y para su velocidad orbital

(54) $$v = \frac{2\pi R}{T} = \sqrt{Rg} \sim 8.231 \text{ metros/seg.}$$

Se puede comparar el valor de T para el satélite que circunvala la Tierra con el período de 27.32 días de la Luna, o sea, el tiempo después del cual ésta regresa a la misma posición entre las estrellas ("mes sideral"). Por la ley de Kepler, la razón de la distancia a entre la Luna y el centro de la Tierra al radio de ésta sería igual a la potencia $\frac{2}{3}$ de la relación de períodos. Esto da para dicha distancia el valor

$$a = \left(\frac{27.32 \times 24}{1.4}\right)^{2/3} R \sim 60R \sim 400{,}000 \text{ Km.},$$

que concuerda bien con el valor promedio real de esa distancia.

c. Movimiento radial — Velocidad de escape

El segundo tipo de movimiento que consideraremos es el de una partícula que se mueve a lo largo de un rayo, por ejemplo, el eje x, desde el centro de atracción. Aquí $y = 0$, $x = r$, de manera que las ecuaciones de movimiento se reducen a

$$\ddot{x} = -\frac{\gamma m_0}{x^2}.$$

Siguiendo el procedimiento general para ecuaciones del tipo $\ddot{s} = f(s)$, multiplicamos ambos miembros de esa ecuación por $\dot{x}$ y obtenemos

$$\dot{x}\ddot{x} = -\gamma m_0 \frac{\dot{x}}{x^2}$$

o bien

$$\frac{d}{dt}\left(\frac{1}{2}\dot{x}^2\right) = \frac{d}{dt}\left(\frac{\gamma m_0}{x}\right).$$

Así, la expresión

$$\frac{1}{2}\dot{x}^2 - \frac{\gamma m_0}{x}$$

tiene un valor constante h durante el movimiento. (Más adelante reconoceremos este hecho como un ejemplo de la ley de la conservación de la energía.) Si introducimos x, en vez de t, como variable independiente, obtenemos

$$\frac{dt}{dx} = \frac{1}{\dot{x}} = \pm\frac{1}{\sqrt{2h + (2\gamma m_0/x)}},$$

que, por integración, conduce a

$$t = t_0 \pm \int_{x_0}^{x} \frac{d\xi}{\sqrt{2h + (2\gamma m_0/\xi)}}.$$

No intentaremos realizar la integración, la cual puede efectuarse fácilmente con ayuda de los métodos desarrollados en el capítulo 3. Para una partícula que se suelta en el tiempo $t_0 = 0$ desde la distancia x_0 con velocidad nula, se tiene $h = -\gamma m_0/x_0$. El tiempo que tarda dicha partícula en caer sobre la partícula atractiva ($x = 0$) es entonces

$$t = \int_0^{x_0} \frac{d\xi}{\sqrt{2\gamma m_0(1/\xi - 1/x_0)}} = \frac{\pi}{2}\sqrt{\frac{x_0^3}{2\gamma m_0}}.$$

Por la ley de Kepler, esto es $\sqrt{1/32}$ veces el tiempo que necesitaría la partícula para dar una vuelta alrededor del centro de atracción a la distancia x_0 [ver Ec. (53), p. 434].

La relación

$$\frac{1}{2}\dot{x}^2 - \frac{\gamma m_0}{x} = h$$

tiene una consecuencia interesante cuando se investigan las condiciones en las que una partícula puede escapar al infinito. Puesto que $\frac{1}{2}\dot{x}^2 \geq 0$, se encuentra que para $x \to \infty$ la constante h debe ser no negativa y, por lo tanto, que $\frac{1}{2}\dot{x}^2 - \gamma m_0/x \geq 0$ durante todo el movimiento. Como caso especial, una partícula que inicia su movimiento desde la distancia $x = a$ con velocidad v puede escapar al infinito sólo si $\frac{1}{2}v^2 - \gamma m_0/a \geq 0$. El valor más bajo posible de la velocidad v que permitirá a la partícula escapar al infinito es entonces $v = \sqrt{2\gamma m_0/a}$. Esta es la *velocidad de escape* v_e. Para una masa que parte de la superficie de la Tierra para escapar al infinito, esto es, escapar de la atracción gravitacional, se tiene $a = R$, $\gamma m_0 = gR^2$, de manera que

$$v_e = \sqrt{2gR} \sim 11{,}300 \text{ metros/seg.}$$

Por consiguiente [cfr. (54), p. 434], la velocidad de escape es justamente $\sqrt{2}$ veces la velocidad requerida para mantener un satélite en órbita circular cerca de la Tierra. Un meteorito que cayese a la Tierra desde el infinito también tendría la velocidad v_e en el momento del impacto, si se desprecian la resistencia del aire y el movimiento de la Tierra en su órbita.

4.9 Trabajo y energía

a. Trabajo realizado por fuerzas durante un movimiento

El *concepto de trabajo* arroja nueva luz sobre las consideraciones de la última sección y sobre muchos otros tópicos en mecánica y física.

Consideremos otra vez una partícula que se mueve sobre una curva bajo la acción de un fuerza que actúa a lo largo de dicha curva, y supongamos que la posición de la partícula está dada por la longitud de arco medida desde cualquier punto inicial fijo. En general, la fuerza que actúa en la dirección del movimiento será a su vez una función de s. Esta función tendrá valores positivos cuando el sentido de la fuerza sea el mismo que el de los valores crecientes de s, y tendrá valores negativos en el caso opuesto.

Si la magnitud de la fuerza es constante a lo largo de toda la trayectoria, entonces se llama *trabajo realizado por la fuerza* al producto de la fuerza por la distancia $(s_1 - s_0)$ recorrida, siendo s_1 el punto final del movimiento y s_0 el punto inicial. Si la fuerza no es constante, definimos el trabajo mediante un proceso de límite. Subdividimos el intervalo de s_0 a s_1 en n subintervalos iguales o diferentes y observamos que si esos subintervalos son pequeños, la fuerza en cada uno de ellos es casi constante; si σ_ν

es un punto escogido arbitrariamente en el ν-ésimo subintervalo, entonces en este subintervalo la fuerza será aproximadamente $f(\sigma_\nu)$. Si la fuerza en el ν-ésimo subintervalo fuese exactamente $f(\sigma_\nu)$, el trabajo realizado por ella sería precisamente

$$\sum_{\nu=1}^{n} f(\sigma_\nu)\,\Delta s_\nu,$$

donde Δs_ν denota, como es costumbre, la longitud del ν-ésimo subintervalo. Si pasamos ahora al límite dejando que n crezca más allá de toda cota en tanto que la longitud del subintervalo mayor tiende a cero, por la definición de integral, la anterior suma tiende entonces a

$$W = \int_{s_0}^{s_1} f(s)\,ds,$$

cantidad que llamamos, en forma natural, el *trabajo* realizado por la fuerza.

Si la fuerza y el movimiento tienen el mismo sentido, el trabajo realizado por ella es positivo; y decimos entonces que *la fuerza efectúa trabajo.* Por otra parte, si la fuerza y el movimiento tienen sentidos opuestos, el trabajo realizado por ella es negativo; y en este caso decimos que *se efectúa trabajo contra la fuerza.*[1]

Si consideramos la coordenada de posición s como una función del tiempo t, de manera que la fuerza $f(s) = p$ sea también una función de t, entonces en una plano con coordenadas rectangulares s y p podemos representar el punto de coordenadas $s = s(t)$ y $p = p(t)$ como una función del tiempo. Este punto describirá una curva que podemos llamar el *diagrama de trabajo* del movimiento. Si se trata de un movimiento periódico, como el de cualquier máquina, entonces después de un cierto tiempo T (un período) el punto móvil $(s(t), p(t))$ regresará a la misma posición; esto es, el diagrama de trabajo será una curva cerrada. En este caso la curva puede consistir simplemente de uno y el mismo arco, recorrido primero hacia adelante y después hacia atrás. Esto ocurre, por ejemplo, en el caso de oscilaciones elásticas. Sin embargo, también es posible que la curva sea de un tipo más general de curva cerrada, limitando un área. Este es el caso, por ejemplo, de las máquinas en que la presión sobre un pistón no es la misma durante la carrera hacia adelante que durante la carrera hacia atrás. El trabajo realizado en un *ciclo,* esto es, en el tiempo T, estará entonces dado simplemente por el negativo del área, con signo negativo, encerrada por el diagrama de trabajo, o sea, en otras palabras, por la integral

[1] Obsérvese que aquí debemos caracterizar cuidadosamente la fuerza de la que hablamos. Por ejemplo, al levantar un peso el trabajo realizado por la fuerza de gravedad es negativo: Se hace trabajo contra la gravedad. Pero desde el punto de vista de la persona que levanta el peso, el trabajo realizado es positivo, pues deberá ejercer una fuerza opuesta a la de gravedad.

$$\int_{t_0}^{t_0+T} p(t)\,\frac{ds}{dt}\,dt,$$

donde el intervalo de tiempo de t_0 a $t_0 + T$ representa exactamente un período del movimiento. Si la frontera del área se recorre en el sentido positivo, el trabajo realizado es negativo; y si se recorre en el sentido negativo, el trabajo realizado será positivo. Si la curva consiste de varias espiras, algunas recorridas positivamente y otras negativamente, el trabajo realizado es la suma de las áreas de las espiras, con su signo cambiado.

Estas consideraciones son ilustradas en la práctica por el *diagrama indicador* de una máquina de vapor antigua. Mediante un dispositivo mecánico diseñado convenientemente, se hace mover un lápiz sobre una hoja de papel; el movimiento horizontal del lápiz relativo al papel es proporcional a la distancia s del pistón a su posición extrema, en tanto que el movimiento vertical es proporcional a la presión del vapor y, por consiguiente, a la fuerza total p del vapor sobre el pistón. Este describe entonces el diagrama de trabajo de la máquina sobre una escala conocida. El área de este diagrama se mide (generalmente usando un planímetro), y así se encuentra el trabajo realizado por el vapor sobre el pistón. Aquí vemos también que nuestra convención para el signo de un área, como se discutió en la p. 382, es definitivamente de interés práctico. Pues ocurre a veces, cuando la máquina funciona baja, que el vapor muy expandido en el extremo de la carrera del pistón tiene una presión menor que la requerida para su expulsión hacia el recorrido de regreso. En el diagrama esto aparece como una espira recorrida positivamente; y la máquina extrae energía del volante en vez de proporcionársela.

b. Trabajo y energía cinética. Conservación de la energía

La ley de movimiento

$$m\ddot{s} = f$$

conduce a una relación fundamental entre los cambios en velocidad durante el movimiento de una partícula a lo largo de una curva y el trabajo realizado por la fuerza f en la dirección del movimiento. Usando el mismo artificio que en los ejemplos anteriores, multiplicamos ambos miembros de la ecuación de movimiento por $\dot{s}$:

$$m\dot{s}\ddot{s} = f(s)\dot{s}.$$

Ahora bien $m\dot{s}\ddot{s} = (d/dt)\frac{1}{2}m\dot{s}^2 = (d/dt)\frac{1}{2}mv^2$, donde $v(t) = \dot{s}$ es la velocidad de la partícula. Integrando ambos lados de la ecuación con respecto a t entre los límites t_0 y t_1, obtenemos

$$\frac{1}{2}mv^2(t_1) - \frac{1}{2}mv^2(t_0) = \int_{t_0}^{t_0} f(s)\,\frac{ds}{dt}dt$$
$$\int_{t_0}^{s_1} f(s)\,ds = W.$$

La cantidad $\frac{1}{2}mv^2$ se llama la *energía cinética, K,* de la partícula. Por lo tanto: *El cambio en energía cinética de una partícula durante su movimiento es igual al trabajo realizado por la fuerza que actúa sobre ella en la dirección del movimiento.*

La cantidad f representa la fuerza que actúa en la dirección del movimiento, o *componente tangencial* de la fuerza. Para una fuerza $\mathbf{F} = (\rho, \sigma)$, su componente en la dirección del movimiento es

$$f = \mathbf{F}\cdot\frac{d\mathbf{R}}{ds} = \rho\frac{dx}{ds} + \sigma\frac{dy}{ds}.$$

Si ρ y σ son funciones conocidas de x e y, y si se sabe que la partícula se mueve sobre una curva $x = x(s)$, $y = y(s)$, entonces f también se vuelve una función conocida de s. En consecuencia, para calcular el trabajo

(55) $$W = \int_{s_0}^{s_1} f(s)\,ds$$

cuando la partícula se mueve de una posición (x_0, y_0) a otra (x_1, y_1), en general tiene que conocerse la trayectoria sobre la cual la partícula se mueve.

En una clase importante de casos el trabajo W depende sólo de las posiciones inicial y final, pudiendo expresarse en la forma

(56) $$W = V(x_0, y_0) - V(x_1, y_1)$$

en la que $V(x,y)$ es una función adecuada que se llama *energía potencial.* La fórmula que expresa que el cambio en energía cinética es igual al trabajo realizado por la fuerza puede entonces escribirse también en la forma

(57) $$\tfrac{1}{2}mv^2(t_1) + V(x_1, y_1) = \tfrac{1}{2}mv^2(t_0) + V(x_0, y_0).$$

Así la cantidad $K + V$, suma de las energías mecánicas cinética y potencial, es decir, la *energía total,* no cambia durante el movimiento. Este es un ejemplo de la ley física general de la *conservación de la energía.*

Una función de energía potencial V puede fácilmente construirse para algunos de los movimientos discutidos antes. Así, para una partícula sujeta a la acción de la gravedad se tiene $\mathbf{F} = (0, -mg)$ y $f = -mg(dy/ds)$. El trabajo realizado por la fuerza de gravedad cuando la partícula pasa de la posición (x_0, y_0) a la posición (x_1, y_1) es entonces

$$W = \int_{s_0}^{s_1} -mg\frac{dy}{ds}\,ds = \int_{y_0}^{y_1} -mg\,dy = mgy_0 - mgy_1.$$

Vemos que W es proporcional a la diferencia de alturas entre las posiciones inicial y final. Como función de energía potencial V, podemos escoger $V = mgy$ (o, más generalmente, $V = mgy + c$, donde c es cualquier constante). La ley de la conservación de la energía establece entonces que la cantidad

$$\tfrac{1}{2}v^2 + gy$$

es constante durante el movimiento. Este hecho ha sido ya observado al investigar el movimiento de una partícula que se desliza hacia abajo sobre una curva (p. 426).

c. La atracción mutua de dos masas

Otro ejemplo de una fuerza con la cual puede asociarse una función de energía potencial V lo proporciona la atracción gravitacional $\mathbf{F}$ que una partícula $P_0 = (x_0, y_0)$ de masa m_0 ejerce sobre otra partícula $P = (x, y)$ de masa m. Aquí es

$$\mathbf{F} = \left[\frac{-\mu(x - x_0)}{r^3}, \frac{-\mu(y - y_0)}{r^3}\right],$$

donde $\mu = \gamma m_0 m$ y $r = \sqrt{(x - x_0)^2 + (y - y_0)^2}$. (Según la ley de Coulomb, el mismo tipo de fórmula describe la interacción de dos cargas eléctricas.)

La fuerza en la dirección del movimiento es entonces

$$f = -\frac{\mu}{r^3}\left[(x - x_0)\frac{dx}{ds} + (y - y_0)\frac{dy}{ds}\right] = -\frac{\mu}{r^2}\frac{dr}{ds} = \frac{d}{ds}\frac{\mu}{r},$$

puesto que

$$(x - x_0)\frac{dx}{ds} + (y - y_0)\frac{dy}{ds} = \frac{1}{2}\frac{d}{ds}[(x - x_0)^2 + (y - y_0)^2]$$

$$= \frac{1}{2}\frac{dr^2}{ds} = r\frac{dr}{ds}.$$

El trabajo realizado por la fuerza de atracción cuando la partícula P se mueve de la posición (x_1, y_1) a la posición (x_2, y_2) es

$$W = \int_{s_1}^{s_2}\left(\frac{d}{ds}\frac{\mu}{r}\right)ds = \frac{\mu}{r_2} - \frac{\mu}{r_1} = V(x_1, y_1) - V(x_2, y_2),$$

donde $V(x, y) = -\mu/r = -\mu/\sqrt{(x - x_0)^2 + (y - y_0)^2}$ es la energía potencial.

Si la partícula se mueve de la posición (x_1, y_1) hasta el infinito (lo que correspondería a $r_2 = \infty$), el trabajo realizado por la fuerza de atracción

es $-\mu/r_1$. El trabajo realizado por una fuerza opuesta que mueva la partícula hasta el infinito tiene el mismo valor numérico pero signo contrario. Por lo tanto, $\mu/r_1 = -V(x_1, y_1)$ es el trabajo que tiene que hacerse *contra* la fuerza de atracción para mover la partícula de la posición (x_1, y_1) hasta el infinito. Esta importante expresión se llama *potencial mutuo* de las dos partículas. Aquí, por consiguiente, el potencial se define como el trabajo requerido para separar completamente las dos masas atractivas, por ejemplo, el trabajo requerido para arrancar por completo un electrón de un átomo (potencial de ionización).

Si la masa atractiva P_0 se considera fija, la ley de la conservación de la energía implica que la partícula atraída P se mueve de manera tal que la expresión

$$\frac{1}{2}v^2 - \frac{\gamma m_0}{r} = h$$

(la energía total por unidad de masa m) tenga un valor constante durante el movimiento. Este hecho ha sido derivado ya para el caso especial de un movimiento puramente radial; y ahora vemos que se cumple para cualquier tipo de movimiento bajo la influencia de la atracción gravitacional. Nuevamente podemos concluir que $h \geq 0$ para una partícula que escapa al infinito; su órbita es entonces no acotada (parábola o hipérbola) en vez de acotada (elipse). La velocidad de escape es

$$v_e = \sqrt{\frac{2\gamma m_0}{r}}$$

la cual corresponde a $h = 0$; es la mínima velocidad que hará que la partícula escape al infinito desde una distancia dada r. No depende de la *dirección* en la cual se suelte la partícula sino sólo de la distancia r al centro de atracción.

d. El estiramiento de un resorte

Como un tercer ejemplo, considérese el trabajo que se hace al estirar un resorte. Con las hipótesis relativas a las propiedades elásticas del resorte que se hicieron en la p. 423, la fuerza actuante es $f = -kx$, donde k es constante. El trabajo que debe hacerse contra esta fuerza para estirar el resorte desde la posición sin deformación $x = 0$ a la posición final $x = x_1$ está dado, en consecuencia, por la integral

$$\int_0^{x_1} kx\,dx = \frac{1}{2}kx_1^2.$$

*e. La carga de un condensador

El concepto de trabajo en otras ramas de la física puede tratarse de manera análoga. Por ejemplo, considérese el problema de cargar un condensador. Si se denota por Q la cantidad de electricidad en el condensador, por C su capacidad y por V la diferencia de potencial (voltaje) entre sus placas, entonces por la física se sabe que $Q = CV$. Además, el trabajo hecho al mover una carga Q a través de una diferencia de potencial V es igual a QV. Puesto que durante el proceso de carga del condensador la diferencia de potencial V no es constante sino que crece con Q, se efectúa un paso al límite de manera análoga a como se hizo en la p. 436, y así se obtiene como expresión para el trabajo realizado al cargar el condensador

$$\int_0^{Q_1} V\,dQ = \frac{1}{C}\int_0^{Q_1} Q\,dQ = \frac{1}{2}\frac{Q_1^2}{C} = \frac{1}{2}Q_1V_1,$$

donde Q_1 es la cantidad total de electricidad que pasa al condensador y V_1 es la diferencia de potencial entre sus placas al final del proceso de carga.

Apéndice

*A.1 Propiedades de la evoluta

En la p. 377 se definió la evoluta E de una curva C como el lugar geométrico de los centros de curvatura de C. Si C está representada por: $x = x(s)$, $y = y(s)$, usando como parámetro la longitud de arco s, entonces el centro de curvatura (ξ, η) del punto de C con parámetro s está dado por [cfr. (17*a*), p. 377]

(58) $$\xi = x - \rho\dot{y}, \qquad \eta = y + \rho\dot{x},$$

donde

$$\kappa = \frac{1}{\rho} = \dot{x}\ddot{y} - \dot{y}\ddot{x}.$$

Las cantidades κ y $|\rho|$ son, respectivamente, la *curvatura* y el *radio de curvatura* de C.

A partir de estas fórmulas pueden deducirse algunas propiedades geométricas interesantes de la evoluta.

Derivando la relación $\dot{x}^2 + \dot{y}^2 = 1$ se llega a $\dot{x}\ddot{x} + \dot{y}\ddot{y} = 0$. Puesto que también $\dot{x}\ddot{y} - \dot{y}\ddot{x} = 1/\rho$, se tiene

(59) $$x = -\frac{1}{\rho}\dot{y}, \qquad \ddot{y} = \frac{1}{\rho}\dot{x}.$$

Derivando con respecto a s las fórmulas (58), resulta

$$\dot{\xi} = \dot{x} - \rho\ddot{y} - \dot{\rho}\dot{y} = -\dot{\rho}\dot{y}, \qquad \dot{\eta} = \dot{y} + \rho\ddot{x} + \dot{\rho}\dot{x} = \dot{\rho}\dot{x},$$

y, en consecuencia,

$$\dot{\xi}\dot{x} + \dot{\eta}\dot{y} = 0.$$

Puesto que los cosenos directores de la normal a la curva están dados por $-\dot{y}$ y $\dot{x}$, *la normal a la curva C es tangente a la evoluta E en el centro de curvatura;* o bien, la tangente a la evoluta es la normal a la curva dada; o, lo que es lo mismo, *la evoluta es la "envolvente" de las normales* (cfr. Fig. A.1).

Además, si se denota por σ la longitud de arco de la evoluta medida desde un punto fijo arbitrario, usando s como parámetro se tendrá

$$\dot{\sigma}^2 = \left(\frac{d\sigma}{ds}\right)^2 = \dot{\xi}^2 + \dot{\eta}^2.$$

Y como $\dot{x}^2 + y = 1$, de las fórmulas (59) se obtiene

$$\dot{\sigma}^2 = \dot{\rho}^2.$$

Si el sentido en que se mide σ se escoge adecuadamente, se sigue que

$$\dot{\sigma} = \dot{\rho},$$

siempre $\rho \neq 0$. La integración da

$$\sigma_1 - \sigma_0 = \rho_1 - \rho_0.$$

Esto es, *la longitud de arco entre dos puntos de la evoluta es igual a la diferencia entre los correspondientes radios de curvatura, suponiendo que $\dot{\rho}$ permanece diferente de cero para el arco en cuestión.*

Esta última condición no es superflua. Pues si $\dot{\rho}$ cambia de signo entonces la fórmula $\dot{\sigma} = \dot{\rho}$ muestra que al pasar al correspondiente punto de la evoluta la longitud de arco σ tiene un máximo o un mínimo; o sea, al pasar a este punto no se continúa considerando σ hacia adelante sino que debe invertirse el sentido en que se mide. Si se desea evitar esta inversión debe cambiarse el signo en la fórmula anterior, esto es, poner $\dot{\sigma} = -\dot{\rho}$ al pasar a tal punto.

También puede notarse que los centros de curvatura que corresponden a máximos o mínimos del radio de curvatura son *cúspides de la evoluta.* [La prueba no se da aquí.] (Ver Figs. A.4, A.6.)

La relación geométrica recién encontrada puede expresarse aún en otra forma: Imaginemos un hilo flexible e inextensible colocado a lo largo de un arco de la evoluta E y estirado de manera que una parte de él se extienda tangencialmente fuera de la curva; si, además, el punto extremo Q del hilo queda inicialmente sobre la curva original C, entonces al desenrollarlo

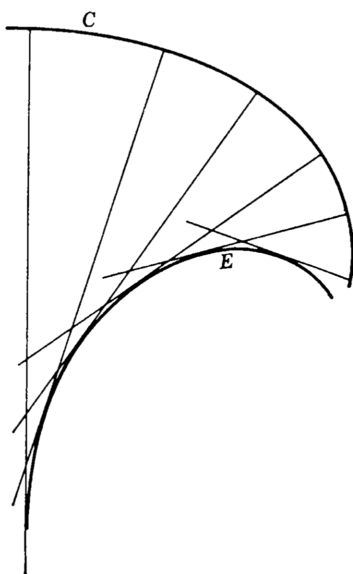


Figura A.1 Evoluta (E).

Q describirá dicha curva C. De aquí proviene el nombre de evoluta *(evolvere,* desenrollar). La curva C se llama una *involuta* de la evoluta E. Por otra parte, podemos comenzar con una curva arbitraria E, y mediante este proceso de desenrollado construir su involuta C. Entonces, en forma inversa se ve que E será la evoluta de C (Fig. A.2).

Para la demostración, considérese la curva E, que ahora es la curva dada, descripta en la forma $\xi = \xi(\sigma), \eta = \eta(\sigma)$, donde las coordenadas rectangulares corrientes se denotan por ξ y η y el parámetro σ es la lon-

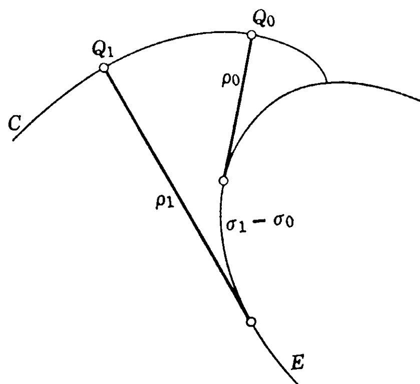


Figura A.2 Construcción de la involuta C de una curva E: $\rho_1 = \rho_0 + \sigma_1 - \sigma_0$, mediante un hilo.

gitud de arco sobre E. El enrollado se hace según lo indica la Fig. A.3. Cuando el hilo está completamente enrollado sobre la evoluta E, el extremo Q de aquél coincide con el punto A de E que corresponde a un arco de longitud a. Si ahora se desenrolla el hilo hasta que sea tangente a la evoluta en el punto P, correspondiente a la longitud de arco $\sigma > a$, la longitud del segmento PQ será $(\sigma - a)$ y sus cosenos directores $-\dot{\xi}$ y $-\dot{\eta}$, donde el punto denota ahora derivación con respecto a σ. Así se obtienen para las coordenadas x, y del punto Q las expresiones

$$(60) \qquad x = \xi - (\sigma - a)\dot{\xi}, \qquad y = \eta - (\sigma - a)\dot{\eta},$$

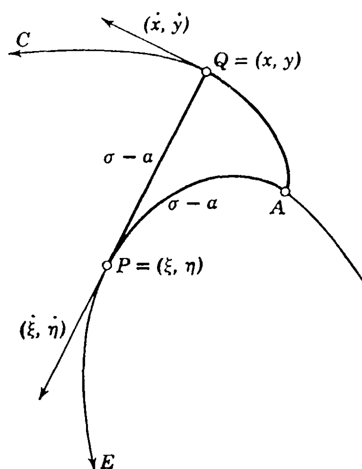


Figura A.3

que son las ecuaciones de la involuta descrita por Q, en términos del parámetro σ. Derivando con respecto a σ resulta

$$(61) \qquad \begin{aligned} \dot{x} &= \dot{\xi} - \dot{\xi} + (a - \sigma)\ddot{\xi} = (a - \sigma)\ddot{\xi}, \\ \dot{y} &= \dot{\eta} - \dot{\eta} + (a - \sigma)\ddot{\eta} = (a - \sigma)\ddot{\eta}. \end{aligned}$$

Puesto que $\dot{\xi}\ddot{\xi} + \dot{\eta}\ddot{\eta} = 0$, se encuentra de inmediato que

$$\dot{\xi}\dot{x} + \dot{\eta}\dot{y} = 0,$$

lo cual muestra que la recta PQ es normal a la involuta C. Por consiguiente, puede afirmarse que las normales a la curva C son tangentes a la curva E. Los cosenos directores de la tangente a E son $\dot{\xi}$ y $\dot{\eta}$, por lo que para los cosenos directores de la tangente a C se encuentran las expresiones

$$(62) \qquad \frac{\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2}} = \dot{\eta}, \qquad \frac{\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}} = -\dot{\xi}.$$

Derivando con respecto a σ la relación $\dot{\xi}\dot{x} + \dot{\eta}\dot{y} = 0$ y substituyendo para $\dot{\xi}$, $\dot{\eta}$, $\ddot{\xi}$, $\ddot{\eta}$ las expresiones que resultan de las ecuaciones (61), (62) anteriores, se llega a

$$0 = \ddot{\xi}\dot{x} + \ddot{\eta}\dot{y} + \dot{\xi}\ddot{x} + \dot{\eta}\ddot{y} = \frac{\dot{x}^2 + \dot{y}^2}{a - \sigma} + \frac{-\ddot{x}\dot{y} + \dot{x}\ddot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}}.$$

Por lo tanto, el radio de curvatura de la curva C correspondiente al punto $Q = (x, y)$ vuelve a ser [ver fórmula (15), p. 373]

$$\rho = \frac{1}{\kappa} = \frac{(\dot{x}^2 + \dot{y}^2)^{3/2}}{\dot{x}\ddot{y} - \dot{y}\ddot{x}} = \sigma - a.$$

Esto representa también la distancia de Q al punto $P = (\xi, \eta)$. Debido a que P también está sobre la normal a C en Q, P resulta ser el centro de curvatura de C correspondiente al punto Q. Así, *cualquier curva E es la evoluta de todas sus involutas.*

Ejemplos. Consideremos la evoluta de la cicloide

$$x = \pi + t + \operatorname{sen} t, \qquad y = -1 - \cos t.$$

Por la Ec. (17) de p. 377, el centro de curvatura (ξ, η) de una curva referida a un parámetro t arbitrario es

$$\xi = x - \dot{y}\frac{\dot{x}^2 + \dot{y}^2}{\dot{x}\ddot{y} - \dot{y}\ddot{x}}, \qquad \eta = y + \dot{x}\frac{\dot{x}^2 + \dot{y}^2}{\dot{x}\ddot{y} - \dot{y}\ddot{x}}.$$

Un breve cálculo da el siguiente resultado para la evoluta de la cicloide

$$\xi = \pi + t - \operatorname{sen} t, \qquad \eta = 1 + \cos t.$$

Si se pone $t = \tau - \pi$, entonces

$$\xi + \pi = \pi + \tau + \operatorname{sen} \tau, \qquad \eta - 2 = -1 - \cos \tau;$$

y estas ecuaciones muestran que la evoluta es a su vez una cicloide similar a la curva original, pudiendo obtenerse a partir de ésta por traslación, como lo indica la Fig. A.4.

Lo anterior nos proporciona un método simple para construir un *péndulo cicloidal* (ver p. 430). Si una masa P se sujeta mediante una cuerda de longitud 4 a una de las cúspides de la evoluta, por efecto de la tensión, parte de la cuerda coincidirá con la evoluta y la parte restante estará a lo largo de una tangente. La masa P es entonces forzada a caer sobre la involuta, o sea, sobre la cicloide original. Bajo la acción de la gravedad P debe efectuar sobre alguna porción de la cicloide un movimiento isócrono, con un período que no depende de la posición en la que P indica su movimiento. (El parámetro t al cual es referida la cicloide *no* corresponde al

tiempo en el movimiento isócrono.) La longitud de la porción recta de la cuerda en un péndulo de este tipo varía durante el movimiento (ver Fig. A.4).

Como un ejemplo adicional se obtiene la ecuación para la involuta de una circunferencia. Se comienza con la circunferencia $\xi = \cos\sigma$, $\eta = -\operatorname{sen}\sigma$ y se desenrolla la tangente como lo indica la Fig. A.5. La involuta de esta curva queda entonces dada en la forma

$$x = \cos\sigma + \sigma\operatorname{sen}\sigma, \qquad y = -\operatorname{sen}\sigma + \sigma\cos\sigma.$$

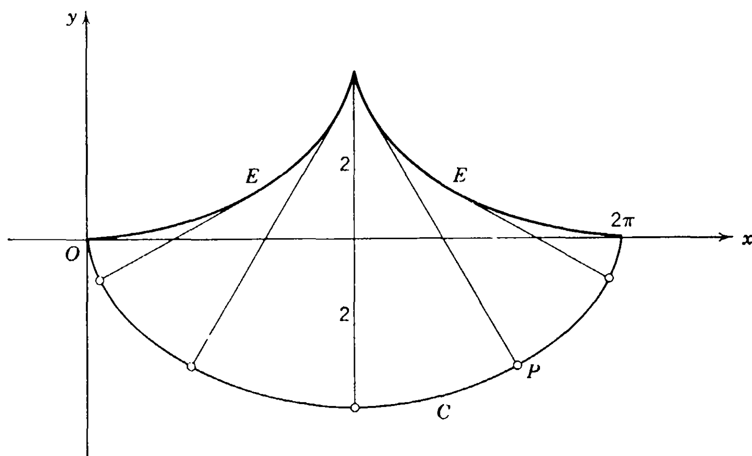


Figura A.4 El péndulo cicloidal.

(usando la ecuación (60) de la p. 445 con $a = 0$).

Por último, se determina la evoluta de la elipse $x = a\cos t$, $y = b\operatorname{sen} t$. De inmediato resulta

$$\xi = x - \dot{y}\frac{\dot{x}^2 + \dot{y}^2}{\dot{x}\ddot{y} - \dot{y}\ddot{x}} = \frac{a^2 - b^2}{a}\cos^3 t.$$

y

$$\eta = y + \dot{x}\frac{\dot{x}^2 + \dot{y}^2}{\dot{x}\ddot{y} - \dot{y}\ddot{x}} = -\frac{a^2 - b^2}{b}\operatorname{sen}^3 t,$$

como representación paramétrica de la evoluta. Si de estas ecuaciones se elimina t de la manera usual, se llega a la ecuación de la evoluta en forma no paramétrica:

$$(a\xi)^{2/3} + (b\eta)^{2/3} = (a^2 - b^2)^{2/3}.$$

Esta curva es una *astroide*. Su gráfica se da en la Fig. A.6. Por medio de las ecuaciones paramétricas podemos fácilmente cerciorarnos de que los centros de curvatura que corresponden a los vértices de la elipse son efectivamente las cúspides de la astroide.

*A.2 Areas limitadas por curvas cerradas. Indices

En la sección 4.2 se representó el área limitada por una curva cerrada $x = x(t)$, $y = y(t)$, $\alpha \leq t \leq \beta$, la cual en ninguna parte se intersecta a sí misma *(curva cerrada simple)*, por medio de la integral

$$A = -\int_{\alpha}^{\beta} y(t)\dot{x}(t)\, dt;$$

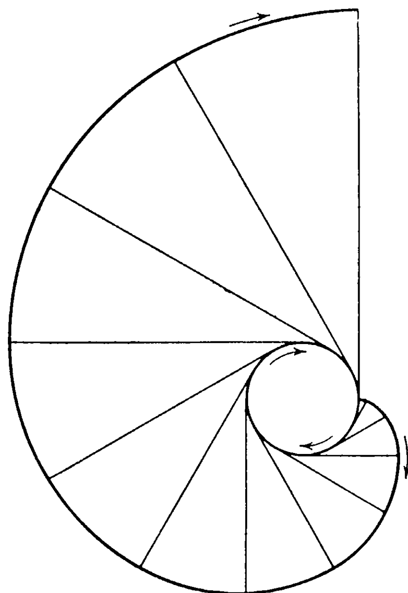


Figura A.5 Involuta de la circunferencia.

el valor que se obtiene es positivo o negativo dependiendo de que la frontera se recorrra en el sentido contra reloj o en el opuesto. Esta fórmula sigue siendo válida como definición de A aun para curvas que presentan autointersecciones. Falta, sin embargo, ver cómo A se relaciona con las áreas en tales casos. Supóngase que la curva C, dada por las ecuaciones $x = x(t)$, $y = y(t)$, se corta a sí misma en un número finito de puntos y de esta manera divide el plano en un número finito de porciones R_1, $R_2, \ldots$. Supóngase además que las derivadas son continuas y que $\dot{x}^2 + \dot{y}^2 \neq 0$, excepto quizá para un número finito de discontinuidades de salto (que pueden corresponder a esquinas o no). Por último, supóngase que la curva tiene un número finito de *rectas de apoyo* $x =$ constante, es decir, rectas verticales que son tangentes a dicha curva o pasan por uno de sus puntos de autointersección.

A cada región R_i se le asigna entonces un entero, el *índice* μ_i, definido de la siguiente manera: Se escoge un punto arbitrario Q de R_i que no esté sobre una recta de apoyo y se traza la semirrecta que se extiende desde Q hacia arriba en la dirección del eje y positivo. Se cuenta el número de veces que la curva C orientada según las t crecientes cruza la semirrecta de derecha a izquierda y se resta del número de veces en que el cruce se

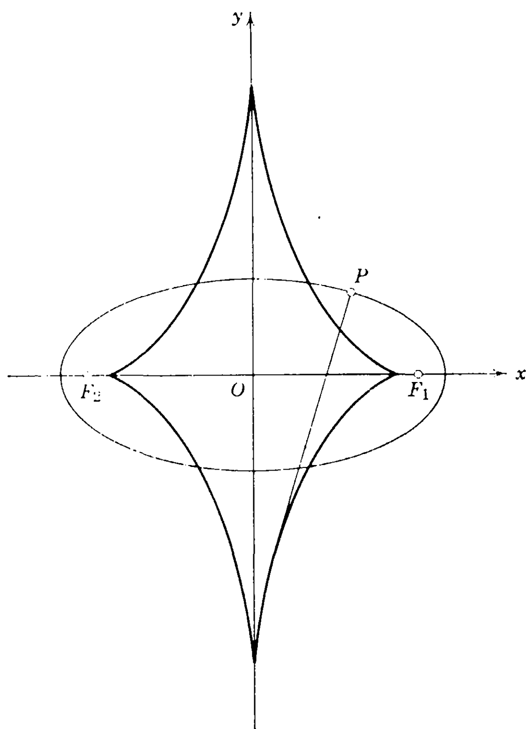


Figura A.6. Evoluta de la elipse.

efectúa de izquierda a derecha; la diferencia es el índice μ_i. Por ejemplo, el interior de la curva representada en la Fig. 4.17, p. 361, tiene el índice $\mu = +1$; y en la Fig. A.7 las regiones $R_1, \ldots, R_5, R_6$ tienen los índices $\mu_1 = -1$, $\mu_2 = -2$, $\mu_3 = -1$, $\mu_4 = 0$, $\mu_5 = 1$ y $\mu_6 = 0$. Este número μ_i efectivamente depende sólo de las regiones R_i y no del punto particular Q escogido en R_i, como puede verse fácilmente de la siguiente manera: Escójase cualquier otro punto Q' en R_i que no esté sobre una recta de apoyo y únase Q con Q' por medio de una línea quebrada contenida completamente en R_i (Fig. A.8). Conforme esta línea se recorre de Q a Q', la diferencia entre el número de cruces de derecha a izquierda y el número de cruces de izquierda a derecha es constante. Esto es así porque entre dos

rectas de apoyo el número de cruces de cada tipo no cambia, mientras que al atravesar la curva una recta de apoyo el número de cruces de ambos tipos sigue siendo el mismo, o bien ambos se incrementan en uno, o bien decrecen en uno; y así, de cualquier manera, la diferencia no se altera. En este proceso una recta de apoyo que encuentra a la curva en varios puntos, digamos $A, B, \ldots, H$, se considera como varias rectas de apoyo diferentes, $FA, FB, \ldots, FH$, donde F es un punto situado verticalmente

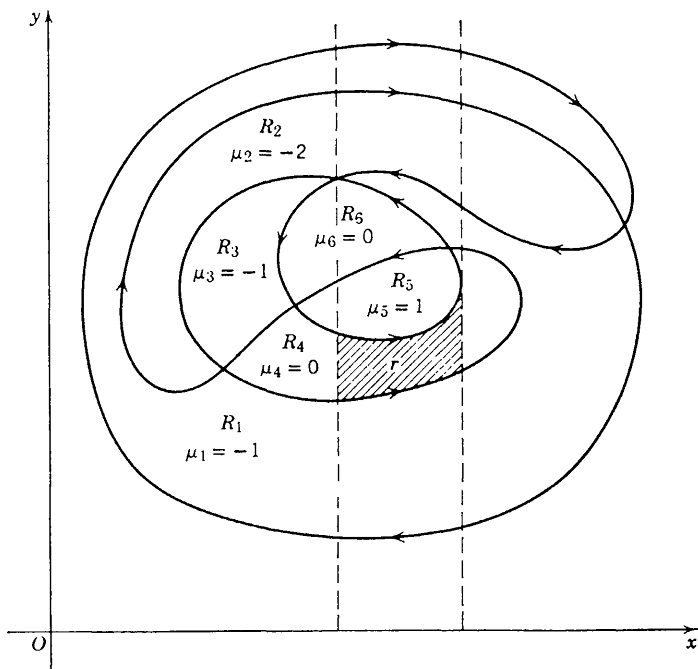


Figura A.7 Indices μ_i de las regiones R_i que forman las curvas cerradas y orientadas. Figura A.8.

por encima de todos los puntos $A, B, \ldots, H$. Nuestro argumento se aplica entonces a cada una de estas rectas. Por lo tanto, el número μ_i tiene el mismo valor independientemente de que para determinarlo se use Q o Q'.

En particular, si la curva no se corta a sí misma, su interior consiste de una única región R cuyo índice es $+1$ o -1 según que la frontera se recorra en el sentido contra reloj o en el opuesto. Para hacer esto evidente, trácese una recta vertical (no de apoyo) que intersecte a la curva; y sobre ella determínese el punto más alto de intersección P con la curva, y escójase un punto Q en R situado debajo de P y tan cercano a él que entre P y Q ningún otro punto de intersección exista. Entonces, por encima de Q se tiene un cruce de la curva que, si ésta se recorre en el sentido contra

reloj, es un cruce de derecha a izquierda, de manera que $\mu = +1$; y en caso contrario $\mu = -1$. Según lo que se vio antes, este mismo valor de μ resulta para cualquier otro punto de R. Para una curva como la considerada y, de hecho, para todas las curvas cerradas, una de las regiones, la "exterior" a la curva, se extiende sin límite en todas direcciones; y vemos de inmediato que esta región tiene índice 0, y será ignorada en lo que sigue. Por lo anterior, la relación entre la integral A y las áreas de las regiones R_i está dada por el siguiente teorema:

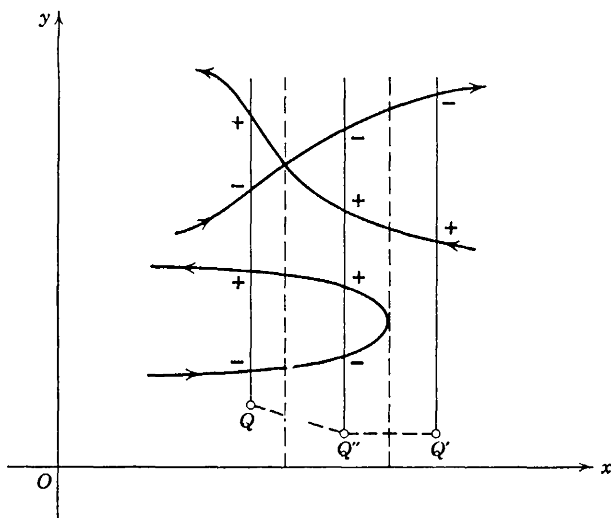


Figura A.8.

TEOREMA. *El valor de la integral* $-\int_\alpha^\beta y\dot{x}\,dt$ *es igual a la suma de las áreas absolutas de las regiones* R_i, *donde cada área* R_i *se cuenta* μ_i *veces. En símbolos:*

$$-\int_\alpha^\beta y\dot{x}\,dt = \sum \mu_i\,|\text{área } R_i|.$$

DEMOSTRACIÓN. La demostración es sencilla. Supongamos, como tenemos derecho a hacerlo, que la curva completa está por encima del eje x. (Agregando una constante a y no se altera el valor de la integral A para una curva cerrada.) Las rectas de apoyo cortan R_i en un número finito de porciones. Sea r una de estas porciones. Entonces, al tomar la integral $-\int y\dot{x}\,dt = -\int y\,dx$ para cada rama univaluada de la función $y = y(x)$ e interpretarla como el área entre la curva y el eje x, encontramos que el valor absoluto del área de r es contada $+1$ veces por cada rama de derecha a izquierda encima de r, y -1 veces por cada rama de izquierda a derecha

encima de r; en total, μ_i veces. Lo mismo es cierto para cada una de las demás porciones de R_i; por lo tanto, R_i se cuenta μ_i veces. Así, la integral sobre la curva completa tiene el valor $\sum \mu_i$ |área R_i|, como se estableció (Ver Fig. A.7). Esta fórmula concuerda con la que se encontró para curvas simples, como se ve a partir de la discusión sobre los valores de μ para tales curvas.

La definición dada para el índice μ_i tiene la desventaja de haber sido formulada en términos de un sistema coordenado particular. De hecho, no obstante, puede demostrarse que el valor de μ_i asignado a cada región R_i es independiente del sistema de coordenadas y depende solamente de la curva. Esto puede verse fácilmente identificando μ_i con el número total ν_i de veces que una porción de la curva, para t creciente de α a β, rodea cualquier punto fijo Q_i de R_i en el sentido contra reloj; esto es, el número de veces que C se *enrolla* alrededor de Q. Probaremos la identidad de μ_i con ν_i.

Sea C dada paramétricamente por $x = x(t)$, $y = y(t)$ donde $\alpha \leq t \leq \beta$. Sea $Q = (\xi, \eta)$ un punto que no está sobre una recta de apoyo de C. Tomemos Q como el origen de un sistema de coordenadas polares r, θ, en el cual

$$r = \sqrt{(x - \xi)^2 + (y - \eta)^2}, \qquad \cos\theta = \frac{x - \xi}{r}, \operatorname{sen}\theta = \frac{y - \eta}{r}.$$

El ángulo polar θ está determinado solamente a menos de múltiplos enteros de 2π; y, sin embargo, θ está determinado unívocamente como una función de t, por su valor θ_0 para $t = \alpha$, si pedimos que $\theta = \theta(t)$ varíe continuamente con t a lo largo de C. En $t = \beta$ el ángulo θ tendrá el valor $\theta(\beta) = \theta_0 + 2\nu\pi$, donde ν es un entero. El número

$$\nu = \frac{1}{2\pi}[\theta(\beta) - \theta(\alpha)] = \frac{1}{2\pi}\int_\alpha^\beta \frac{d\theta}{dt}dt = \frac{1}{2\pi}\int_C d\theta$$

representa el número de veces que la curva orientada C rodea a Q.

Así, la curva C cruza la semirrecta vertical que pasa por Q, para aquellos valores de t en los que la expresión $(1/2\pi)[\theta(t) - \pi/2]$ tiene un valor entero n. Considérense para un n fijo los valores t en el intervalo paramétrico para los cuales $(1/2\pi)(\theta - \pi/2) = n$. Sean σ_n y τ_n los números de tales valores t para los cuales $d\theta/dt > 0$ y $d\theta/dt < 0$, respectivamente. Es obvio que el *índice* en el punto Q es

$$\mu = \sum_n \sigma_n - \sum_n \tau_n = \sum_n (\sigma_n - \tau_n).$$

Por otra parte, $\sigma_n - \tau_n$ puede solamente tener uno de los valores 1, 0, -1, pues la gráfica de $\theta(t)$ en el plano θ, t debe cruzar la recta $\theta = \pi/2 + 2n\pi$ alternadamente desde arriba y desde abajo. Precisamente, se tiene $\sigma_n - \tau_n =$

$\operatorname{sign}[\theta(\beta) - \theta(\alpha)]$ si $\pi/2 + 2n\pi$ está entre $\theta(\alpha)$ y $\theta(\beta)$; de lo contrario, es $\sigma_n - \tau_n = 0$.

En consecuencia, μ es igual al número de valores de la forma $\pi/2 + 2n\pi$, con un n entero que está entre $\theta(\alpha)$ y $\theta(\beta)$, tomado con el signo de $\theta(\beta) - \theta(\alpha)$; o sea, μ es igual a ν.

Puesto que $\theta = \arctan[(y - \eta)/(x - \xi)]$, se tiene

$$\frac{d\theta}{dt} = \frac{\dot{y}(x - \xi) - \dot{x}(y - \eta)}{(x - \xi)^2 + (y - \eta)^2}$$

Esto conduce a la siguiente representación integral para el índice de la curva C cerrada, orientada, con respecto al punto (ξ, η):

$$\mu = \frac{1}{2\pi}\int_\alpha^\beta \frac{\dot{y}(x - \xi) - \dot{x}(y - \eta)}{(x - \xi)^2 + (y - \eta)^2}\,dt,$$

que puede escribirse en forma simple (ver pp. 384-385) sin referencia explícita al parámetro t, como

$$\mu = \frac{1}{2\pi}\int_C \frac{(x - \xi)\,dy - (y - \eta)\,dx}{(x - \xi)^2 + (y - \eta)^2}.$$

La característica notable de estos resultados es que el entero μ, o el ν, que describe una relación topológica entre el punto Q y la curva C, puede determinarse analíticamente a partir de la representación paramétrica de C, evaluando una integral.

PROBLEMAS

SECCION 4.1*c*, página 328

1. Bosquejar la hipocicloide para $a = 4c$ (astroide) y encontrar su ecuación no paramétrica.

2. Demostrar que si c/a es racional la hipocicloide general se cierra después de que el círculo móvil ha girado un número entero de veces, en tanto que si c/a es irracional la curva encuentra a la circunferencia del círculo fijo en una infinidad de puntos y nunca se cierra.

3. Obtener la representación paramétrica

$$x = at - b \operatorname{sen} t, \qquad y = a - b \cos t$$

para la trocoide ordinaria, esto es, para la trayectoria de un punto P sujeto a un disco de radio a que rueda a lo largo de una recta, siendo b la distancia de P al centro del disco (ver Fig. 4.7).

4. Encontrar las ecuaciones paramétricas de la curva $x^3 + y^3 = 3axy$ (la hoja de Descartes), escogiendo como parámetro t la tangente del ángulo entre el eje x y el rayo que va del origen al punto (x, y).

SECCION 4.1*e*, página 361

1. El ángulo α entre dos curvas en un punto de intersección se define como el ángulo entre sus tangentes en ese punto. Encontrar una fórmula para $\cos\alpha$ en términos de las representaciones paramétricas de las curvas.

2. Sean $x = f(t)$ e $y = g(t)$. Encontrar fórmulas para d^2y/dx^2 y d^3y/dx^3 en términos de las derivadas con respecto al parámetro t.

3. Encontrar la fórmula en coordenadas polares para el ángulo α entre dos curvas $r = f(\theta)$ y $r = g(\theta)$.

4. Encontrar las ecuaciones de las curvas que en todas partes se intersectan bajo el mismo ángulo α con las rectas que pasan por el origen.

5. Demostrar lo siguiente: si $x = f(t)$ e $y = g(t)$ son continuas en el intervalo cerrado $[a, b]$ y derivables en el intervalo abierto (a, b), con $x'^2 + y'^2 > 0$, entonces existe al menos un punto sobre el arco abierto

$$x = f(t), \qquad y = g(t), \qquad (a < t < b),$$

en que la tangente es paralela a la cuerda que une los puntos extremos.

6. Sea P el punto de un círculo que describe una cicloide conforme el círculo rueda sobre una recta dada. Sea, además, Q el punto de contacto del círculo con la recta. Demostrar que en cualquier instante la normal a la cicloide en P pasa por Q. ¿Qué propiedad análoga se tiene para la tangente en P?

7. Probar que la longitud del segmento de la tangente a la astroide

$$x = 4c \cos^3 \theta, \qquad y = 4c \operatorname{sen}^3 \theta,$$

cortado por los ejes de coordenadas es constante.

*8. Demostrar que las dos familias de elipses e hipérbolas, $(0 < a < b)$,

$$\frac{x^2}{a^2 - \lambda^2} + \frac{y^2}{b^2 - \lambda^2} = 1, \quad \text{para } 0 < \lambda < a,$$

$$\frac{x^2}{a^2 - \tau^2} + \frac{y^2}{b^2 - \tau^2} = 1, \quad \text{para } a < \tau < b,$$

son cofocales (esto es, tienen los mismos focos) y se intersectan a ángulos rectos.

9. (*a*) Demostrar que para la elipse el ángulo entre los dos rayos que van de cada foco a un punto sobre la curva es bisectado por la normal en ese punto.
(*b*) Demostrar que para la hipérbola ese ángulo es bisectado por la tangente.

SECCION 4.1*f*, página 366

1. Probar que la curva definida por

$$y = \begin{cases} x^2 \operatorname{sen} \dfrac{1}{x}, & 0 < x \leq 1 \\ 0, & x = 0 \end{cases}$$

tiene longitud finita, pero que la curva continua definida por

$$y = \begin{cases} x \operatorname{sen} \dfrac{1}{x}, & 0 < x \leq 1 \\ 0, & x = 0 \end{cases}$$

no es rectificable.

2. Probar que si la función f está definida y es monótona en el intervalo cerrado $[a, b]$, entonces el arco definido por

$$y = f(x), \qquad (a \leq x \leq b),$$

es rectificable.

SECCION 4.1*g*, página 370

1. Una ***integral elíptica de segunda clase*** tiene la forma

$$\int_0^{\phi} \sqrt{1 - k^2 \operatorname{sen}^2 \theta}\, d\theta.$$

(*a*) Demostrar que la longitud de arco de la elipse $x = a \cos \theta$, $y = b \operatorname{sen} \theta$ puede expresarse en términos de una integral elíptica de segunda clase,

(*b*) Hacer lo mismo para lo trocoide

$$x = at - b \operatorname{sen} t, \qquad y = a - b \cos t.$$

*(*c*) Demostrar que la longitud de arco de la hipérbola puede expresarse en términos de integrales elípticas de primera y segunda clases.

SECCION 4.1*h*, página 372

1. Sea P un punto del círculo que al rodar genera una cicloide, y sea Q el punto más bajo de dicho círculo en cualquier instante dado.

Demostrar que Q bisecta el segmento que une a P con el centro de la circunferencia osculatriz de la cicloide en P.

2. Encuéntrese el centro de curvatura para $y = x^2$ cuando $x = 0$. Determinar el punto de intersección de las normales a la curva para $x = 0$ y $x = \varepsilon$. Calcular la distancia de la intersección al centro de curvatura. Sugerir una definición alternativa para el centro de curvatura. Demostrar que esta definición es equivalente a la definición dada en el texto.

3. Considérese la cuestión de si la circunferencia osculatriz cruza la curva en el punto de contacto.

*4. Demostrar que la circunferencia de curvatura de la curva C en un punto P es el límite de las circunferencias que pasan por los tres puntos P, P_1, P_2, conforme P_1 y P_2 se acercan a P.

5. Sea $r = f(\theta)$ la ecuación de una curva en coordenadas polares. Probar que la curvatura está dada por la fórmula

$$k = \frac{2r'^2 - rr'' + r^2}{(r'^2 + r^2)^{3/2}},$$

donde

$$r' = \frac{df}{d\theta}, \qquad r'' = \frac{d^2 f}{d\theta^2}.$$

6. La curva para la cual el segmento de tangente definido por el punto de contacto y la intersección con el eje y tiene longitud siempre igual a 1, se llama ***tractriz***. Encontrar su ecuación. Demostrar que el radio de curvatura en cualquier punto de la curva es inversamente proporcional a la longitud del segmento de normal definido por el punto sobre la curva y la intersección con el eje y. Calcular la longitud de arco de la tractriz y encontrar las ecuaciones paramétricas en términos de la longitud de arco.

7. Sea $x = x(t)$, $y = y(t)$ una curva cerrada. Una longitud constante p es medida sobre la normal desde el punto de su intersección con la curva. El extremo de este segmento describe una curva que se llama "curva paralela" a la original. Encontrar el área, la longitud de arco y el radio de curvatura de la "curva paralela".

8. Demostrar que las únicas curvas cuya curvatura es una constante fija k son circunferencias de radio $1/k$.

*9. Si la curvatura de una curva en el plano xy es una función monótona de la longitud de arco, demostrar que la curva no es cerrada y que no tiene puntos dobles.

SECCION 4.1*i*, página 378

1. Demostrar que la expresión para la curvatura de una curva $x = x(t)$, $y = y(t)$ no es alterada por una rotación de ejes y tampoco por un cambio de parámetro dado por $t = \phi(\tau)$, donde $\phi'(\tau) > 0$.

SECCION 4.3*d*, página 411

1. Demostrar que si la aceleración es siempre perpendicular a la velocidad, la velocidad (escalar) es constante.

2. El vector velocidad, considerado como un vector de posición, describe una curva conocida como odógrafa. Para una partícula que se mueve sobre una curva cerrada, investigar si la odógrafa de la partícula puede o no ser una recta.

3. Suponiendo que el círculo que rueda se mueve con velocidad constante, encontrar la velocidad y la aceleración del punto P que genera la cicloide.

4. Sea A un punto fijo en el plano y supóngase que el vector aceleración de un punto móvil P está siempre dirigido hacia A y es proporcional a $1/|AP|^2$. Demostrar que la odógrafa es una circunferencia (ver problema 2).

5. Sea A un punto fijo sobre un círculo. Sea P el punto del círculo móvil que se mueve de manera que el vector aceleración apunta hacia A. Probar que la aceleración es proporcional a $|AP|^{-5}$.

SECCION 4.5, página 420

1. Una partícula que se mueve sobre una recta está sujeta a una resistencia que produce una retardación ku^3, donde u es la velocidad y k una constante. Encontrar expresiones para la velocidad (u) y el tiempo (t) en términos de s, la distancia desde la posición inicial, y v_0, la velocidad inicial.

2. Una partícula de masa unitaria se mueve sobre el eje x sujeta a una fuerza $f(x) = -\operatorname{sen} x$.
(*a*) Determinar el movimiento de la partícula si en el tiempo $t = 0$ se encuentra en $x = 0$ y tiene velocidad $v_0 = 2$. Demostrar que cuando $t \to \infty$ la partícula se aproxima a una posición límite, y encontrar esta posición.
(*b*) Si las condiciones son las mismas, excepto que v_0 puede tener cualquier valor, demostrar que para $v_0 > 2$ el punto se mueve hacia una distancia infinita cuando $t \to \infty$, y que para $v_0 < 2$ el punto oscila en torno al origen.

3. Escójase un sistema de ejes con origen en el centro de la Tierra, cuyo radio se denota por R. De acuerdo con la ley de la gravitación de Newton, una partícula de masa unitaria que esté sobre el eje y será atraída por la Tierra con una fuerza $-\mu M/y^2$, donde μ es la "constante gravitacional" y M es la masa terrestre.
(*a*) Determinar el movimiento de la partícula después de que ésta se suelta desde el punto $y_0(> R)$; esto es, si en el tiempo $t = 0$ está en el punto $y = y_0$ y tiene velocidad $v_0 = 0$.
(*b*) Encontrar la velocidad de la partícula en el momento de su impacto con la Tierra.
(*c*) Usando el resultado de (*b*), calcular la velocidad de una partícula que cae a la Tierra desde el infinito.[1]

*4. Una partícula en reposo sobre el punto más alto de una circunferencia es ligeramente perturbada y se desliza hacia abajo por acción de la gravedad. ¿En qué punto dejará de moverse sobre la circunferencia?

[1] Esta es la misma que la mínima velocidad con la que debería dispararse un proyectil para que dejase la Tierra sin regresar jamás.

*5. Una partícula de masa m se mueve sobre una elipse $r = k/(1 - e\cos\theta)$. La fuerza sobre la partícula es cm/r^2, dirigida hacia el origen. Describir el movimiento de la partícula, encontrar su período y demostrar que el radio vector de la partícula barre áreas iguales en tiempos iguales ("Ley de las Areas", de Kepler, para el movimiento planetario).

SECCION 4A.1, página 442

1. Demostrar que la evoluta de una epicicloide (ejemplo, p. 348) es otra epicicloide similar a la original que puede obtenerse de ésta por rotación y contracción.

2. Demostrar que la evoluta de una hipocicloide (ejemplo, p. 348) es otra hipocicloide que puede obtenerse de la primera por rotación y dilatación.

5

Desarrollo de Taylor

5.1 Introducción: Series de potencias

Fue un gran triunfo en los primeros años del Cálculo el descubrimiento hecho por Newton y otros de que muchas funciones conocidas podían expresarse como "polinomios de orden infinito" o "series de potencias", con coeficientes calculados mediante leyes elegantes y claras. Las series geométricas para $1/(1-x)$ y para $1/(1+x^2)$,

$$(1) \qquad \frac{1}{1-x} = 1 + x + x^2 + \cdots + x^n + \cdots$$

$$(1a) \qquad \frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + \cdots + (-1)^n x^{2n} + \cdots,$$

válidas en el intervalo abierto $|x| < 1$, son prototipos (ver capítulo 1, página 90).

Desarrollos análogos de la forma

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n + \cdots$$
$$= \sum_{\nu=0}^{\infty} a_\nu x^\nu,$$

con coeficientes numéricos a_ν, serán deducidos en este capítulo para muchas otras funciones.

Los siguientes son ejemplos notables:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots;$$

$$\operatorname{sen} x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - + \cdots + \frac{(-1)^n x^{2n+1}}{(2n+1)!} + \cdots;$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{2!} + \cdots + \frac{(-1)^n x^{2n}}{(2n)!} + \cdots$$

Estos desarrollos en series son válidos para todo x.

Teorema general del binomio de Newton. El desarrollo

$$(1+x)^\alpha = 1 + \frac{\alpha}{1!}x + \frac{\alpha(\alpha-1)}{2!}x^2 + \cdots$$

$$= \sum_{\nu=0}^{\infty} \binom{\alpha}{\nu} x^\nu$$

es válido para $|x| < 1$ y cualquier exponente α.

Para explicar el significado preciso de tales desarrollos, considérese el polinomio de orden n formado por la suma de los $n + 1$ primeros términos de la serie, esto es, la *n-ésima "suma parcial"*,

$$S_n = \sum_{\nu=0}^{n} a_\nu x^\nu.$$

La fórmula

$$f(x) = \sum_{\nu=0}^{\infty} a_\nu x^\nu, \quad \text{para } |x| < a$$

significa entonces: Para $n \to \infty$ la sucesión S_n tiende al valor de la función $f(x)$ en cada punto del intervalo $|x| < a$. Se dice entonces que la serie infinita *converge* a $f(x)$ en el intervalo $|x| < a$. La diferencia

$$R_n(x) = f(x) - S_n(x),$$

el *"residuo"* de la serie, mide la precisión con la cual el polinomio $S_n(x)$ aproxima a la función $f(x)$ en x. Por ejemplo,

$$(1b) \qquad \frac{1}{1-x} = 1 + x + x^2 + \cdots + x^n + R_n(x),$$

donde el residuo $R_n(x) = x^{n+1}/(1-x)$ tiende a cero para $|x| < 1$ conforme n crece; y resulta así la serie geométrica infinita $\sum_{\nu=0}^{\infty} x^\nu = 1/(1-x)$. Encontrar en casos específicos estimaciones sencillas para R_n que sean fácilmente utilizables es una tarea de importancia tanto teórica como práctica.

En este capítulo estamos interesados en tales desarrollos para una amplia clase de funciones, incluyendo todas las funciones trascendentales "elementales". Es notable el hecho de que en estos desarrollos de funciones trascendentes los coeficientes sean expresiones simples en términos de enteros. El tratamiento de estos desarrollos se hará mediante el *teorema de Taylor.* Más adelante, en el capítulo 7, se discutirá un enfoque diferente por medio del estudio directo de las series de potencias.

Debe hacerse notar que a menudo, como ocurre para la serie geométrica de la Ec. ($1a$), el desarrollo infinito no es válido fuera de algún intervalo de x en el caso de la serie geométrica, el intervalo $x^2 < 1$), aunque la función representada por la serie esté bien definida fuera de este intervalo.

5.2 Desarrollo del logaritmo y de la tangente inversa

a. El logaritmo

Como ejemplos sencillos se deducen primero los desarrollos para las funciones logaritmo y tangente inversa usando integración y a partir de la serie geométrica

$$\frac{1}{1-t} = 1 + t + t^2 + \cdots + t^{n-1} + r_n(t)$$

con $r_n(t) = t^n/(1-t)$.

El integrando en la fórmula

$$-\log(1-x) = \int_0^x \frac{dt}{1-t}$$

se reemplaza por la suma anterior y se realiza la integración término a término, obteniéndose para $x < 1$

$$-\log(1-x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \cdots + \frac{x^n}{n} + R_n(x),$$

donde el residuo es

$$R_n(x) = \int_0^x r_n\,dt = \int_0^x \frac{t^n}{1-t}\,dt.$$

Por tanto, para cualquier entero positivo n la función $-\log(1-x)$ es aproximada por el polinomio de grado n

$$x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots + \frac{x^n}{n},$$

donde el residuo R_n indica el "error" de esta aproximación.

Para evaluar la exactitud de dicha aproximación se estima el residuo R_n. Si primero se supone que $-1 \leq x \leq 0$, entonces en todo el intervalo de integración el valor absoluto del integrando $t^n/(1-t)$ nunca excede a $|t^n| = (-1)^n t^n$. Así,

$$|R_n| \leq \left| \int_0^x t^n\,dt \right| = \frac{|x|^{n+1}}{n+1}.$$

por tanto, para todo valor de x en el intervalo *cerrado* $-1 \leq x \leq 0$, incluyendo $x = -1$, este residuo puede hacerse tan pequeño como se quiera con sólo escoger n suficientemente grande (ver p. 84). Cuando $x > 0$, el punto extremo $x = 1$ debe omitirse y x tiene que restringirse al intervalo $0 \leq x < 1$. Aquí el integrando no cambia de signo y su valor absoluto no supera a $t^n/(1 - x)$. Así se obtiene para $0 \leq x < 1$ la estimación

$$|R_n| \leq \frac{1}{1-x}\int t^n\,dt = \frac{x^{n+1}}{(1-x)(n+1)}.$$

Otra vez, por tanto, si x es fijo el residuo es arbitrariamente pequeño para n suficientemente grande. Por supuesto, la estimación anterior no tiene significado para $x = 1$.

En resumen,

$$(2) \qquad \log(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots - \frac{x^n}{n} - R_n$$

donde el residuo R_n tiende a cero conforme n crece, suponiendo que x está en el intervalo semiabierto $-1 \leq x < 1$.

De hecho, este razonamiento establece una estimación "uniforme" para el residuo, independiente de x y válida para todo x en el intervalo $-1 \leq x \leq 1 - h$, donde h es cualquier número tal que $0 < h \leq 1$, a saber, $|R_n| \leq 1/[(n+1)h]$.

El hecho de que el residuo R_n tienda a cero en el intervalo semiabierto $-1 \leq x < 1$ queda expresado diciendo que en este intervalo la función logaritmo está dada por la serie infinita [1]

$$(3) \qquad \log(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \cdots$$

Si en esta fórmula se inserta el valor particular $x = -1$, se obtiene una fórmula notable:

$$(4) \qquad \log 2 = 1 - \tfrac{1}{2} + \tfrac{1}{3} - \tfrac{1}{4} + - \cdots.$$

Esta es una de las relaciones cuyo descubrimiento produjo una honda impresión en los iniciadores del cálculo.

Para el intervalo abierto $-1 < x < 1$ sólo se necesita reemplazar x por $-x$ en (2) para obtener

$$(2a) \qquad \log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + - \cdots + (-1)^{n-1}\frac{x^n}{n} - R_n',$$

[1] Se deja como ejercicio al lector verificar que para todos los valores de x tales que $|x| > 1$ el residuo no solamente deja de tender a cero sino que, de hecho, R_n crece más allá de toda cota al crecer n; de manera que para tales valores de x el polinomio ya no es una buena aproximación del logaritmo, sino que empeora al aumentar n.

donde

$$R_n'(x) = \int_0^{x} \frac{t^n\,dt}{1-t} = (-1)^{n+1}\int_0^{x} \frac{t^n\,dt}{1+t}.$$

Tomando n par y restando (2) de (2a) se tiene

$$\frac{1}{2}\log\left(\frac{1+x}{1-x}\right) = \text{ar}\tanh x = x + \frac{x^3}{3} + \frac{x^5}{5} + \cdots + \frac{x^{n-1}}{n-1} + \bar{R}_n,$$

donde el residuo $\bar{R}_n$ está dado por

$$\bar{R}_n = \frac{1}{2}(R_n - R_n') = \int^{x} \frac{t^n}{1-t^2}\,dt,$$

y arc tanh x se define como en la p. 254.

Si se observa que $1/(1-t^2) \leq 1/(1-x^2)$, se encuentra, mediante una estimación elemental de la integral, que

$$|\bar{R}_n| \leq \frac{|x^{n-1}|}{n+1}\cdot\frac{1}{1-x^2};$$

de esta manera el residuo $\bar{R}_n$ tiende a cero conforme n crece, lo cual se expresa escribiendo el desarrollo como una serie infinita:

$$(5) \qquad \frac{1}{2}\log\frac{1+x}{1-x} = \text{ar}\tanh x = x + \frac{x^3}{3} + \frac{x^5}{5} + \frac{x^7}{7} + \cdots,$$

para todos los valores de x con $|x| < 1$. Es de notar, este resultado también podría derivarse directamente integrando la serie geométrica para $1/(1-x^2)$. Esta fórmula presenta la ventaja de que al recorrer x el intervalo de -1 a 1, la expresión $(1+x)/(1-x)$ corre sobre todos los números positivos. Así, si el valor de x se escoge en forma adecuada, la serie nos permite calcular el valor del logaritmo de cualquier número positivo con un error no mayor que el valor estimado arriba como cota para $\bar{R}_n$.

b. La tangente inversa

La tangente inversa puede tratarse de manera similar a la usada para el logaritmo, partiendo con la fórmula

$$\frac{1}{1+t^2} = 1 - t^2 + t^4 - + \cdots + (-1)^{n-1}t^{2n-2} + r_n$$

donde es ahora $r_n = (-1)^n \dfrac{t^{2n}}{1+t^2}$.

Por integración [ver Ec. (14), p. 284] se obtiene

$$\operatorname{arc\,tan} x = x - \frac{x^3}{3} + \frac{x^5}{5} - + \cdots + (-1)^{n-1}\frac{x^{2n-1}}{2n-1} + R_n,$$

$$R_n = (-1)^n \int_0^x \frac{t^{2n}}{1+t^2}\,dt;$$

y se ve de inmediato que en el intervalo cerrado $-1 \leq x \leq 1$ el residuo R_n tiende a cero conforme n crece, pues

$$|R_n| \leq \int_0^{|x|} t^{2n}\,dt = \frac{|x|^{2n+1}}{2n+1}.$$

De la fórmula para el residuo puede también fácilmente demostrarse que para $|x| > 1$ el valor absoluto del residuo crece más allá de toda cota conforme n crece.

De esta manera se ha deducido la serie infinita

(6) $$\operatorname{arc\,tan} x = x - \frac{x^3}{3} + \frac{x^5}{5} - + \cdots + (-1)^{n-1}\frac{x^{2n-1}}{2n-1} + - \cdots,$$

válida para el intervalo cerrado $|x| \leq 1$. Puesto que $\operatorname{arc\,tan} 1 = \pi/4$, se obtiene para $x = 1$ la serie de *Leibnitz-Gregory*

(7) $$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - + \cdots,$$

una expresión tan notable como la encontrada antes para log 2.

5.3 Teorema de Taylor

Taylor, alumno de Newton, observó que el desarrollo elemental de polinomios se prestaba a una amplia generalización para funciones no polinomiales suponiendo que estas funciones fuesen suficientemente derivables y con dominios restringidos adecuadamente.

a. Representación de Taylor para polinomios

Es ésta una fórmula algebraica completamente elemental relativa a un polinomio en x de orden n, por ejemplo

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n.$$

Si se reemplaza x por $a + h = b$ y se desarrolla cada término en potencias de h resulta inmediatamente una representación en la forma

(8) $$f(a+h) = c_0 + c_1 h + c_2 h^2 + \cdots + c_n h^n.$$

La fórmula de Taylor está dada por la relación

$$(8a) \qquad c_\nu = \frac{1}{\nu!} f^{(\nu)}(a),$$

para el coeficiente c_ν en términos de f y sus derivadas en $x = a$. Para demostrar esto, consideremos la cantidad $h = b - a$ como la variable independiente y apliquemos la regla de la cadena, según la cual derivar con respecto a h es lo mismo que derivar con respecto a $b = a + h$. De esta manera, derivando una y otra vez la fórmula (8) con respecto a h y substituyendo $h = 0$ se obtienen sucesivamente los resultados

$$c_0 = f(a),\ c_1 = f'(a),\ \ldots,\ \nu!\, c_\nu = f^{(\nu)}(a),$$

y, por consiguiente, la fórmula de Taylor para polinomios:

$$(9) \qquad f(a + h) = f(a) + hf'(a) + \frac{h^2}{2} f''(a) + \cdots + \frac{h^n}{n!} f^{(n)}(a).$$

La derivada $(n + 1)$-ésima se anula para un polinomio de grado n, de manera que nuestra fórmula (9) termina naturalmente.

La fórmula (9), tal como se la establece, no es más que un elemental reordenamiento algebraico de un polinomio en potencias de $a + h$ para dar un polinomio en potencias de h.

b. Fórmula de Taylor para funciones no polinomiales

Newton y sus discípulos inmediatos aplicaron audazmente la fórmula (9) a funciones no polinomiales, para las cuales el desarrollo no termina automáticamente en el n-ésimo término. En vez de cortar el desarrollo permitieron que n tendiera a infinito, procedimiento que para muchas funciones especiales importantes será justificado posteriormente.

Suponiendo que la función f es derivable por lo menos n veces en un intervalo que contiene los puntos a y $a + h$, ya no podemos escribir $f(a + h)$ como una expresión de la fórmula de (9) consistente de un número finito de potencias de h, sino que debe tomarse en cuenta la discrepancia introduciendo un "residuo" adicional R_n, lo cual se hace escribiendo formalmente

$$(10) \qquad f(b) = f(a + h) = f(a) + hf'(a) + \cdots + \frac{h^n}{n!} f^n(a) + R_n.$$

De hecho (10), no es sino la definición del *término residual* de corrección R_n e indica la esperanza de que R_n se vuelva pequeño y, más aún, tienda a cero para $n \to \infty$. Si, efectivamente, el residuo tiende a cero, entonces la fórmula (10) en el límite $n \to \infty$ conduce al desarrollo

$$(11) \qquad f(a+h) = f(a) + hf'(a) + \cdots + \frac{h^n}{n!} f^{(n)}(a) + \cdots$$

para $f(x)$ como una *serie infinita de potencias* de h.

El problema crucial, que mucho supera en dificultad al de las manipulaciones algebraicas de la sección 5.3*a*, es entonces el de encontrar estimaciones para el residuo R_n de manera que la exactitud de la representación mediante el *polinomio de Taylor* de orden n en h,

$$(12) \qquad T_n(h) = \sum_0^n \frac{f^{(\nu)}(a)}{\nu!} h^\nu;$$

y el paso al límite para $n \to \infty$, puedan explorarse rigurosamente. El polinomio de Taylor $T_n(h)$ es una aproximación a $f(a+h)$ en el sentido de que en $h = 0$ las funciones T_n y f, así como sus derivadas hasta el orden n, coinciden, de manera que la diferencia $R_n = f - T_n$ se anula en $x = a$ junto con las n primeras derivadas.

5.4 Expresiones y estimaciones para el residuo

a. Expresiones de Cauchy y de Lagrange

Una representación directa del residuo R_n que permite hacer estimaciones de su valor absoluto, $|R_n|$, es lo que constituye el núcleo del teorema de Taylor. Los resultados se obtienen fácilmente sobre la base del teorema del valor medio del Cálculo. Están, además, relacionados con la *aproximación lineal de funciones mediante diferenciales* (ver p. 200).

Primeramente examinemos de nuevo esta aproximación.

La definición de derivada en un punto a establece simplemente que $f(a+h) = f(a) + hf'(a) + h\varepsilon$, donde $\varepsilon \to 0$ para $h \to 0$. Podemos lograr una aproximación algo más fina averiguando si ε es, en efecto, de orden por lo menos igual al de h, siempre que tanto f' como f'' existan y sean continuas en nuestro intervalo J. La estimación se obtiene si escribimos nuevamente $a + h = b$, introducimos un residuo R mediante

$$(13) \qquad f(b) = f(a) + (b-a)f'(a) + R,$$

y consideramos ahora b fijo y el punto inicial a variable. Esta ecuación define R como una función de a en el intervalo J. Derivando entonces con respecto a la variable a se obtiene cero para el miembro izquierdo, pues $f(b)$ es constante, y la regla para derivar un producto muestra que

$$0 = f'(a) - f'(a) + (b-a)f''(a) + R'(a)$$

y, por tanto,

$$-R'(a) = (b - a)f''(a). \tag{14}$$

Ahora bien, para $a = b$ tenemos obviamente $R(b) = 0$. Por el teorema del valor medio del Cálculo se tiene $[R(a) - R(b)]/(b - a) = -R'(\xi)$, donde ξ es algún valor entre a y b; y, debido a que $R(b) = 0$, concluimos por consiguiente que $R(a) = -(b - a)R'(\xi) = -hR'(\xi)$. Entonces, por (14) se tiene $R'(\xi) = -(b - \xi)f''(\xi)$, y, por tanto $|R'(\xi)| < h\,|f''(\xi)|$ puesto que $|b - \xi| \leq h$. Como $|f''(\xi)|$ es acotado en un intervalo en torno a a, obtenemos finalmente una estimación que muestra que el residuo o "error" R_n es pequeño al menos hasta el segundo orden en h:

$$|R(a)| < h^2\,|f''(\xi)|. \tag{15}$$

Del caso especial $n = 1$ pasamos ahora al de cualquier orden n. La caracterización directa del residuo R_n se logra mediante el mismo artificio que para $n = 1$. Suponemos que a y $b = a + h$ son puntos de un intervalo J en el cual $f(x)$ está definida y tiene derivadas continuas hasta de orden $n + 1$. Consideramos a como la variable independiente y mantenemos fijo el punto extremo b. En la fórmula (10), p. 465, que define $R_n(a)$, escribimos $b - a$ en vez de h. Al derivar y tomar en cuenta que $f(b)$ es constante, se encuentra, por la regla del producto, que casi todos los términos se cancelan, quedando la fórmula

$$0 = \frac{(b - a)^n}{n!} f^{(n+1)}(a) + R_n'(a) \tag{16}$$

para todo valor a en el intervalo. Puesto que para $a = b$ el residuo R_n es cero, esta expresión directa para su derivada como función de a caracteriza completamente a R_n como una integral $\int_b^a R_n'(t)\,dt = -\int_a^b R_n'(t)\,dt$; o sea:

$$R_n(a) = \int_a^b \frac{(b - t)^n}{n!} f^{n+1}(t)\,dt. \tag{17}$$

Esta es una *representación integral exacta para el residuo.*

Una estimación para R_n análoga a la obtenida antes para $n = 1$ se desprende directamente del teorema del valor medio del Cálculo, aplicado a (16):

$$\frac{R_n(a) - R_n(b)}{b - a} = \frac{R_n(a)}{b - a} = -R_n'(\xi) = \frac{(b - \xi)^n}{n!} f^{n+1}(\xi);$$

o sea

$$R_n(a) = \frac{(b - a)(b - \xi)^n}{n!} f^{n+1}(\xi), \tag{18}$$

donde ξ es un valor no especificado que se elige adecuadamente entre a y b. La misma estimación puede también obtenerse aplicando a la expresión (17) el teorema del valor medio del cálculo integral (capítulo 2, p. 163).

Forma de Cauchy para el residuo. Si se define $\xi = a + \theta h = a + \theta(b - a)$, se obtiene la *fórmula de Cauchy* para el residuo en la fórmula de Taylor (10):

$$R_n(a) = \frac{h^{n+1}}{n!}(1 - \theta)^n f^{(n+1)}(a + \theta h), \tag{19}$$

donde θ es una cantidad entre 0 y 1 no especificada.

Puede también aplicarse a la integral (17) para R_n el *teorema generalizado del valor medio* del cálculo integral (ver p. 164), tomando para la "función peso" $p(t)$ la expresión $p(t) = (b - t)^n$, que no cambia de signo en el intervalo de integración.[1] Entonces,

$$R_n = \frac{1}{n!} f^{(n+1)}(\xi) \int_a^b (b - t)^n \, dt = \frac{(b - a)^{n+1}}{(n + 1)!} f^{(n+1)}(\xi). \tag{20}$$

Forma de Lagrange para el residuo. Haciendo de nuevo $\xi = a + \theta h$ se obtiene la forma de *Lagrange* para el residuo,

$$R_n(a) = \frac{h^{n+1}}{(n + 1)!} f^{(n+1)}(a + \theta h), \tag{21}$$

con una cantidad adecuada θ tal que $0 \leq \theta \leq 1$. La forma de Lagrange es particularmente sugestiva, y, por lo tanto, más comúnmente aplicada, puesto que hace que el residuo en la fórmula

$$f(a + h) = f(a) + \frac{h}{1!} f'(a) + \frac{h^2}{2!} f''(a) + \cdots + \frac{h^n}{n!} f^{(n)}(a) + R_n = P_n(h) + R_n \tag{22}$$

se parezca al término $h^{n+1} f^{(n+1)}(a)/(n + 1)!$ que aparecería en el desarrollo (22) para un orden mayor, con sólo reemplazar el argumento a por el valor intermedio $a + \theta h$.

Para una función f cuya derivada $f^{(n+1)}$ es continua en un intervalo cerrado conteniendo al punto a, la cantidad $|f^{(n+1)}(\xi)|$ tiene una cota fija M. Puesto que entonces

$$|R_n| \leq \frac{h^{n+1}}{(n + 1)!} M,$$

[1] El teorema generalizado del valor medio fue demostrado para el caso de una $p(t)$ positiva; pero vale igualmente cuando $p(t)$ es negativa en todo el intervalo de integración.

el polinomio de Taylor $P_n(h)$ da para un n fijo una aproximación a la función $f(a + h)$ con un error de orden al menos $n + 1$ en h.

Nuestro interés se enfocará principalmente sobre la cuestión de si el residuo R_n tiende a cero conforme n crece; y si éste es el caso, decimos que la función se ha desarrollado en una serie de Taylor infinita

$$(23) \qquad f(a + h) = f(a) + \frac{h}{1!}f'(a) + \frac{h^2}{2!}f''(a) + \frac{h^3}{3!}f'''(a) + \cdots;$$

en particular, si primero se hace $a = 0$ y después se escribe x en lugar de h, se obtiene la "serie de potencias"

$$f(x) = f(0) + \frac{x}{1!}f'(0) + \frac{x^2}{2!}f''(0) + \cdots$$

En la sección 5.5 se discutirán algunos ejemplos.

Para las aplicaciones, igualmente importante es el desarrollo finito de Taylor para un n fijo, con el término del residuo. Si se hace h tender a cero en esta fórmula, según la terminología del capítulo 3, p. 273, los diferentes términos de la serie tienden a cero con diferentes órdenes de magnitud en h. La expresión $f(a)$ representa el término de orden cero en la serie de Taylor; la expresión $hf'(a)$, el término de primer orden; la expresión $h^2f''(a)/2!$ el término de segundo orden, etc. De la forma para el residuo se observa que al desarrollar una función hasta el término de orden n se comete un error que tiende a cero con orden $(n + 1)$ conforme h tiende a cero. Mientras más próximo a a esté $a + h$, mejor será la representación de la función $f(a + h)$ mediante el polinomio de aproximación $P_n(h)$. En los casos de mayor interés la aproximación en la inmediata vecindad del punto x puede mejorarse aumentando el valor de n.

b. Una derivación alternativa de la fórmula de Taylor

La representación integral (17) para el residuo R_n en el teorema de Taylor se basó en la fórmula (16) para $R_n'(a)$. Debido a la importancia del teorema, damos aquí una versión diferente de su derivación, que lleva directamente a expresar R_n mediante repetidas *integraciones por partes* comenzando con la fórmula

$$(24) \qquad f(b) - f(a) = \int_a^b f'(t)\, dt.$$

Para transformar (24) usando sucesivamente integración por partes, se introducen las funciones

$$\phi_1(t), \phi_2(t), \ldots, \phi_\nu(t), \ldots,$$

por medio de las relaciones

(25) $$\phi_0(t) = 1, \qquad \phi_\nu'(t) = \phi_{\nu-1}(t),$$

y de las condiciones

(26) $$\phi_\nu(b) = 0, \qquad \text{para} \quad \nu \geq 1,$$

donde b se considera como un parámetro fijo. Claramente, las condiciones (25) y (26) determinan en forma sucesiva todas las $\phi_\nu(t)$. Como puede verificarse de inmediato, las $\phi_\nu(t)$ son precisamente los polinomios

$$\phi_\nu(t) = \frac{(t-b)^\nu}{\nu!}.$$

De paso observamos que las funciones ϕ_ν se obtienen una de otra mediante integraciones sucesivas, dejando indeterminadas las constantes de integración; por consiguiente, las condiciones de definición (25) podrían satisfacerse también por otras funciones que cumplieran condiciones colaterales diferentes a las (26) (ver p. 210).

Puesto que $\phi_\nu(a) = (-1)^\nu(b-a)^\nu/\nu!$ y $\phi_\nu(b) = 0$, se obtiene

$$f(b) - f(a) = \int_a^b \phi_0 f'\,dt = \int_a^b \phi_1' f'\,dt = \phi_1 f'\Big|_a^b - \int_a^b \phi_1 f''\,dt;$$

e integrando por partes otra vez el último término, resulta

$$\begin{aligned} f(b) - f(a) &= (b-a)f'(a) - \int_a^b \phi_2' f''\,dt \\ &= (b-a)f'(a) + \frac{(b-a)^2}{2!}f''(a) + \int_a^b \phi_2 f'''\,dt, \end{aligned}$$

y repitiendo n veces el proceso,

$$\begin{aligned} f(b) - f(a) &= (b-a)f'(a) + \frac{(b-a)^2}{2!}f''(a) + \cdots + \frac{(b-a)^n}{n!}f^{(n)}(a) \\ &\quad + \int_a^b (-1)^n \phi_n(t) f^{(n+1)}(t)\,dt \\ &= (b-a)f'(a) + \frac{(b-a)^2}{2!}f''(a) + \cdots + \frac{(b-a)^n}{n!}f^{(n)}(a) + R_n, \end{aligned}$$

donde, por definición de ϕ_n,

$$R_n = \int_a^b f^{(n+1)}(t)\,\frac{(b-t)^n}{n!}\,dt.$$

Hemos así demostrado nuevamente el

TEOREMA DE TAYLOR. *Si una función $f(t)$ tiene derivadas continuas hasta de orden $n+1$ en un intervalo cerrado que contiene los puntos a y b, entonces:*

$$f(b) = f(a) + (b - a)f'(a) + \cdots + \frac{(b-a)^n}{n!} f^{(n)}(a) + R_n,$$

donde el residuo R_n, que depende de n, a y b, está dado por

$$R_n = \frac{1}{n!}\int_a^b (b - t)^n f^{(n+1)}(t)\, dt. \tag{27}$$

Mediante cambios de notación se obtienen expresiones ligeramente diferentes para la fórmula de Taylor. Así, reemplazando a por x y b por $x + h$ se tiene

$$f(x + h) = f(x) + hf'(x) + \cdots + \frac{h^n}{n!} f^{(n)}(x) + R_n, \tag{27a}$$

con

$$R_n = \frac{1}{n!}\int_x^{x+h} (x + h - t)^n f^{(n+1)}(t)\, dt;$$

y si $t = x + \tau$,

$$R_n = \frac{1}{n!}\int_0^h (h - \tau)^n f^{(n+1)}(x + \tau)\, d\tau. \tag{27b}$$

Haciendo $x = 0$ y escribiendo x en lugar de h, se obtiene [1]

$$f(x) = f(0) + \frac{x}{1!} f'(0) + \frac{x^2}{2!} f''(0) + \cdots + \frac{x^n}{n!} f^{(n)}(0) + R_n \tag{27c}$$

con el residuo

$$R_n = \frac{1}{n!}\int_0^x (x - t)^n f^{(n+1)}(t)\, dt.$$

Aplicando el teorema del valor medio del cálculo integral, o su forma generalizada, a la integral anterior, se llega, respectivamente, a la fórmula de Cauchy

$$R_n = \frac{(1 - \theta)^n}{n!} x^{n+1} f^{(n+1)}(\theta x)$$

y a la fórmula de Lagrange

$$R_n = \frac{x^{n+1}}{(n + 1)!} f^{(n+1)}(\theta x)$$

[1] Este caso especial del teorema, algunas veces se lo llama, sin justificación histórica, teorema de Maclaurin. El teorema general de Taylor fue publicado en 1715; el resultado especial de Maclaurin, en 1742.

para el residuo, tal como se vio antes (p. 467). Aquí θ es un número no especificado pero elegido adecuadamente y tal que $0 \leq \theta \leq 1$ (no necesariamente el mismo en ambas fórmulas).

Como un ejercicio, el lector podría construir funciones ϕ_ν que satisfagan (25) pero para las cuales las condiciones (26) se reemplazan por las relaciones

$$\int_0^1 \phi_\nu(t)\, dt = 0$$

para $\nu \geq 1$ (ver capítulo 8, Apéndice A).

5.5 Desarrollos de funciones elementales

Los resultados generales anteriores nos permiten desarrollar en serie de Taylor las funciones elementales más sencillas. En el capítulo 7 se discutirán los desarrollos para otras funciones.

a. La función exponencial

En primer lugar desarrollamos la función exponencial, $f(x) = e^x$. En este caso todas las derivadas son idénticas a $f(x)$ y tienen el valor 1 para $x = 0$. La forma de Lagrange para el residuo (p. 468, ecuación (21) conduce de inmediato a la fórmula:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \frac{x^{n+1}}{(n+1)!} e^{\theta x}; \qquad 0 < \theta < 1.$$

Si ahora dejamos que n crezca más allá de toda cota, el residuo R_n tiende a cero para cualquier valor fijo de x. Para demostrar esto, nótese primero que $e^{\theta x} \leq e^{|x|}$, puesto que e^x es una función monótona creciente. Sea m cualquier entero mayor que $2|x|$. Entonces para todo $k > m$ se cumple $|x|/k < \frac{1}{2}$ y

$$\left|\frac{x^{n+1}}{(n+1)!}\right| = \frac{|x^m|}{m!} \cdot \frac{|x|}{m+1} \cdots \frac{|x|}{n+1}$$
$$\leq \frac{|x^m|}{m!} \cdot \frac{1}{2^{n+1-m}} \leq \frac{|2x|^m}{m!} \cdot \frac{1}{2^n},$$

de manera que

$$|R_n| \leq \frac{|2x|^m}{m!} \cdot e^{|x|} \cdot \frac{1}{2^n}.$$

Puesto que los dos primeros factores de la derecha son independientes de n, en tanto que $1/2^n \to 0$ para $n \to \infty$, nuestra afirmación queda demostrada.

La función e^x se representa, por consiguiente, mediante la serie infinita

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$
$$= \sum_{\nu=0}^{\infty} \frac{x^\nu}{\nu!}.$$

Este desarrollo es *válido para todos los valores de* x. En particular, para $x = 1$ se obtiene de nuevo la serie infinita que sirvió para definir el número e en el capítulo 1 (ver p. 100).

Por supuesto, para cálculos numéricos debe usarse la forma del teorema de Taylor con residuo. Por ejemplo, para $x = 1$ (compárese con el cálculo similar de la p. 101) se tiene

$$e = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} + \frac{e^\theta}{(n+1)!}.$$

Si se desea calcular e con un error máximo de 1/10,000, lo único que se necesita es tomar n tan grande que el residuo sea menor que 1/10,000 y, puesto que este residuo es ciertamente menor[1] que $3/(n+1)!$, basta escoger $n = 7$, ya que $8! > 30{,}000$. De esta manera se obtiene el valor aproximado $e = 2.71825$, cuyo error es menor que 0.0001.

b. Desarrollos de sen x, cos x, senh x, cosh x

Para las funciones sen x, cos x, senh x, cosh x se encuentran las siguientes fórmulas·

$f(x) =$	$\operatorname{sen} x$	$\cos x$	$\operatorname{senh} x$	$\cosh x$
$f'(x) =$	$\cos x$	$-\operatorname{sen} x$	$\cosh x$	$\operatorname{senh} x$
$f''(x) =$	$-\operatorname{sen} x$	$-\cos x$	$\operatorname{senh} x$	$\cosh x$
$f'''(x) =$	$-\cos x$	$\operatorname{sen} x$	$\cosh x$	$\operatorname{senh} x$
$f^{(4)}(x) =$	$\operatorname{sen} x$	$\cos x$	$\operatorname{senh} x$	$\cosh x$

Así, en los polinomios de aproximación para sen x y senh x los coeficientes de las potencias pares de x se anularán, en tanto que para cos x y cosh x los coeficientes que se anulan son los de las potencias impares.

---

[1] Aquí se ha usado el hecho de que $e < 3$. Esto surge (ver p. 101) de nuestra serie para e; pues siempre es cierto que $1/n! \leq 1/2^{n-1}$, y por consiguiente

$$e < 1 + 1 + \tfrac{1}{2} + \tfrac{1}{4} + \ldots = \frac{1}{1 - \tfrac{1}{2}} = 3.$$

Cuando se usa la forma de Lagrange para el residuo, (21), p. 468, las series de Taylor para nuestras funciones resultan ser:

$$\operatorname{sen} x = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots + \frac{(-1)^n x^{2n+1}}{(2_n + 1)!}$$
$$+ \frac{(-1)^{n+1} x^{2n+3} \cos(\theta x)}{(2_n + 3)!};$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - + \cdots + \frac{(-1)^n x^{2n}}{(2n)!}$$
$$+ \frac{(-1)^{n+1} x^{2n+2} \cos(\theta x)}{(2n + 2)!},$$

$$\operatorname{senh} x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots + \frac{x^{2n+1}}{(2n + 1)!}$$
$$+ \frac{x^{2n+3} \cosh(\theta x)}{(2n + 3)!},$$

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots + \frac{x^{2n}}{(2n)!}$$
$$+ \frac{x^{2n+2} \cosh(\theta x)}{(2n + 2)!}$$

donde, por supuesto, en cada una de nuestras fórmulas θ denota un número diferente en el intervalo $0 \leq \theta \leq 1$, número que, además, depende de n y de x. Puesto que en cada una de estas fórmulas el residuo tiende a cero conforme n crece, lo cual puede verse por exactamente el mismo argumento que el usado para e^x, pueden hacerse las aproximaciones tan precisas como se desee. Se obtienen así las cuatro series infinitas, válidas para cualquier valor de x,

$$\operatorname{sen} x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - + \cdots = \sum_{\nu=0}^{\infty} \frac{(-1)^\nu x^{2\nu+1}}{(2\nu + 1)!},$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - + \cdots = \sum_{\nu=0}^{\infty} \frac{(-1)^\nu x^{2\nu}}{(2\nu)!},$$

$$\operatorname{senh} x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots = \sum_{\nu=0}^{\infty} \frac{x^{2\nu+1}}{(2\nu + 1)!},$$

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots = \sum_{\nu=0}^{\infty} \frac{x^{2\nu}}{(2\nu)!}.$$

Las últimas dos pueden también obtenerse formalmente a partir de la serie para e^x de acuerdo con las definiciones de las funciones hiperbólicas (ver p. 250).

c. La serie binomial

Volvemos a las series de Taylor para las funciones log $(1 + x)$ y arc tan x, ya tratadas directamente en la sección 5.2. Sin embargo, trataremos la generalización del teorema binomial para exponentes arbitrarios, lo cual constituyó uno de los descubrimientos matemáticos de Newton más espectaculares. Deseamos desarrollar la función $f(x) = (1 + x)^\alpha$ en una serie de Taylor para el caso en que $x > -1$ y α es un número arbitrario, positivo o negativo, racional o irracional. Se escoge la función $(1 + x)^\alpha$ en vez de x^α porque para esta última no es cierto que todas las derivadas sean continuas para $x = 0$, excepto en el caso trivial de valores de α enteros no negativos. Primero calculamos las derivadas de $f(x)$, obteniendo

$$f'(x) = \alpha(1 + x)^{\alpha-1},$$
$$f''(x) = \alpha(\alpha - 1)(1 + x)^{\alpha-2}, \ldots,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$f^{(\nu)}(x) = \alpha(\alpha - 1)\cdots(\alpha - \nu + 1)(1 + x)^{\alpha-\nu}$$

En particular, para $x = 0$ resulta

$$f'(0) = \alpha,$$
$$f''(0) = \alpha(\alpha - 1), \ldots,$$
$$f^{(\nu)}(0) = \alpha(\alpha - 1)\cdots(\alpha - \nu + 1).$$

El teorema de Taylor establece entonces que

$$(1 + x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha - 1)}{2!}x^2 + \cdots + \frac{\alpha(\alpha - 1)(\alpha - 2)\cdots(\alpha - n + 1)}{n!}x^n + R_n.$$

Convergencia

Debemos todavía discutir el residuo. Este problema no es muy difícil pero, sin embargo, no es tan simple como los casos tratados anteriormente. Obtendremos una estimación para el residuo directamente y también como un caso especial del resultado general de la sección A.4. Esto nos permitirá concluir que, siempre que $|x| < 1$, el residuo R_n para el desarrollo binomial tiende a cero. Así, la expresión $(1 + x)^\alpha$ podrá desarrollarse en una serie binomial infinita

$$(1 + x)^\alpha = 1 + \frac{\alpha}{1!}x + \frac{\alpha(\alpha - 1)}{2!}x^2 + \cdots$$
$$= \sum_{\nu=0}^{\infty}\binom{\alpha}{\nu}x^\nu,$$

donde, por brevedad, se han introducido los coeficientes binomiales generales

$$\binom{\alpha}{\nu} = \frac{\alpha(\alpha-1)\cdots(\alpha-\nu+1)}{\nu!} \qquad (\text{para } \nu > 0),$$

$$\binom{\alpha}{0} = 1.$$

*Para ver directamente que $R_n \to 0$ para $n \to \infty$ en el caso en que $-1 < x < 1$, se hace uso de la forma de Cauchy para el residuo, (19), p. 468:

$$\begin{aligned} R_n &= \frac{(1-\theta)^n}{n!} x^{n+1} f^{(n+1)}(\theta x) \\ &= \frac{(1-\theta)^n}{n!} \alpha(\alpha-1)(\alpha-2)\cdots(\alpha-n) x^{n+1}(1+\theta x)^{\alpha-n-1} \end{aligned}$$

$(0 \leq \theta \leq 1)$. Puesto que $|x| < 1$, se tiene $0 \leq (1-\theta)/(1+\theta x) \leq 1$, de manera que

$$|R_n| \leq (1+\theta x)^{\alpha-1} |\alpha x| \left|\left(1-\frac{\alpha}{1}\right)x\right| \left|\left(1-\frac{\alpha}{2}\right)x\right| \cdots \left|\left(1-\frac{\alpha}{n}\right)x\right|$$

Existe un número q tal que $|x| < q < 1$. Entonces es obvio que también

$$\left|\left(1-\frac{\alpha}{m}\right)x\right| < q$$

para todo m suficientemente grande (digamos, $m > N$). Así, para $n > N$ se cumple

$$|R_n| \leq (1+\theta x)^{\alpha-1} |\alpha| (1+|\alpha|)^N q^{n-N}.$$

El factor $(1+x\theta)^{\alpha-1}$ es acotado (por $2^{\alpha-1}$ si $\alpha \geq 1$, y por $(1-q)^{\alpha-1}$ si $(\alpha < 1)$, de manera que, claramente, $R_n \to 0$.

Una fórmula un poco más general da una expresión para $(a+b)^\alpha$. Sólo tiene que factorizarse a^α y aplicarse el desarrollo binomial con $x = b/a$ para obtener, con $a > 0$ y $|b| < a$,

$$\begin{aligned} (a+b)^\alpha = a^\alpha\left(1+\frac{b}{a}\right)^\alpha &= a^\alpha\left(1+\alpha\frac{b}{a}+\frac{\alpha(\alpha-1)}{1\cdot 2}\left(\frac{b}{a}\right)^2+\cdots\right) \\ &= a^\alpha + \frac{\alpha}{1}a^{\alpha-1}b + \frac{\alpha(\alpha-1)}{1\cdot 2}a^{\alpha-2}b^2 + \cdots. \end{aligned}$$

5.6 Aplicaciones geométricas

El comportamiento de una función $f(x)$ en una vecindad del punto $x = a$, o el comportamiento de una cierta curva en la vecindad de uno de

sus puntos, puede describirse en detalles mediante el teorema de Taylor, ya que éste nos permite descomponer el incremento de la función, conforme se recorre una vecindad del punto $x = a + h$, en una suma de cantidades de primer orden, segundo orden, etc., en h.

a. Contactos entre curvas

Contacto de orden superior

Si en un punto $x = a$ dos curvas, $y = f(x)$ e $y = g(x)$, se intersectan y tienen una tangente común, se dice que las curvas *se* tocan una con otra o que tienen *contacto de primer orden.* En este caso los desarrollos de Taylor para las funciones $f(a + h)$ y $g(a + h)$ tienen los mismos términos de orden cero y de primer orden en h. Si, además, en el punto $x = a$ las segundas derivadas de $f(x)$ y $g(x)$ son iguales entre sí, se dice que las curvas tienen *contacto de segundo orden.* Entonces, los términos de segundo orden en los desarrollos de Taylor de f y de g también serán idénticos. Si se supone que ambas funciones tienen derivadas continuas por lo menos hasta de tercer orden, entonces la diferencia

$$D(x) = f(x) - g(x)$$

puede expresarse en la forma

$$\begin{aligned} D(a + h) &= f(a + h) - g(a + h) \\ &= \frac{h^3}{3!} D'''(a + \theta h) = \frac{h^3}{3!} F(h), \end{aligned}$$

donde la expresión $F(h)$ tiende a $f'''(a) - g'''(a)$ conforme h tiende a cero. Por lo tanto, la diferencia $D(a + h)$ se hace nula al menos de tercer orden con h.

Puede proseguirse de esta manera y considerarse el caso general en que las series de Taylor para $f(x)$ y $g(x)$ coincidan hasta los términos de n-ésimo orden: esto es,

$$f(a) = g(a),\ f'(a) = g'(a), \ldots, f^{(n)}(a) = g^{(n)}(a).$$

Se supone que las $(n + 1)$-ésimas derivadas son continuas. Bajo estas condiciones se dice que las curvas definidas por nuestras dos funciones tienen *contacto* de *n-ésimo orden* en el punto $x = a$. La diferencia de las dos funciones es entonces de la forma

$$\begin{aligned} D(a + h) = f(a + h) - g(a + h) &= \frac{h^{n+1}}{(n + 1)!} D^{(n+1)}(a + \theta h) \\ &= \frac{h^{n+1}}{(n + 1)!} F(h), \end{aligned}$$

donde la cantidad $F(h) = D^{(n+1)}(a + \theta h)$ tiende a $f^{(n+1)}(a) - g^{(n+1)}(a)$ conforme h tiende a cero, puesto que $0 \leq \theta \leq 1$. De esta fórmula se ve que en el punto de contacto. la diferencia $f(x) - g(x)$ se anula con, por lo menos, $(n + 1)$-ésimo orden.

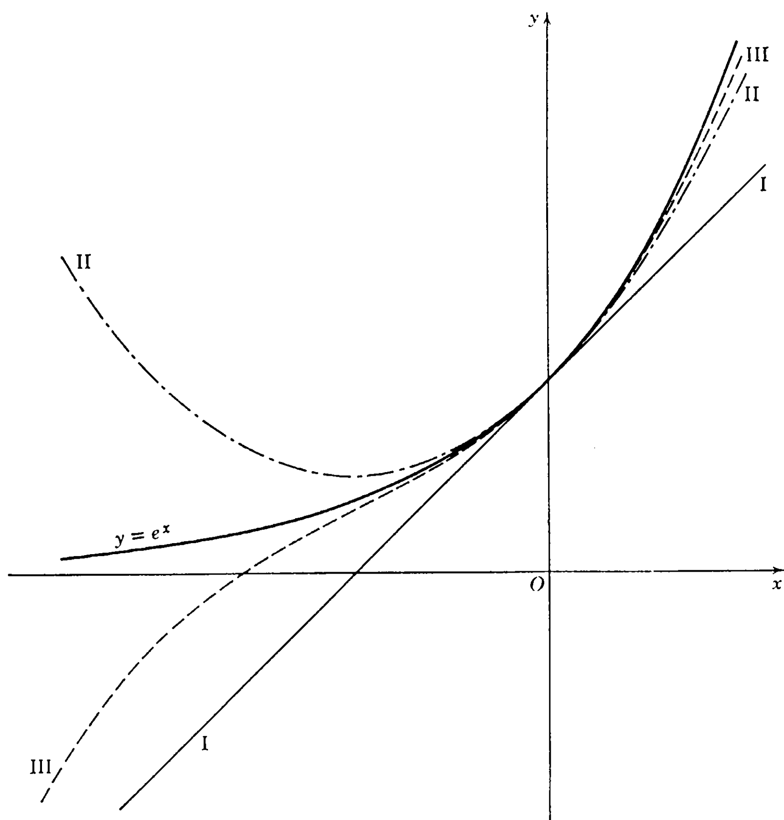


Figura 5.1 Parábolas osculadoras de e^x.

Los polinomios de Taylor definidos por

$$P_n(x) = f(a) + \frac{x-a}{1!}f'(a) + \cdots + \frac{(x-a)^n}{n!}f^{(n)}(a)$$

se caracterizan geométricamente como las "parábolas" de orden n que tienen contacto del mayor orden posible con la gráfica de la función en el punto dado. Por tanto, estas parábolas se llaman a veces *parábolas osculadoras*. (Solamente para $n = 2$ son estas curvas "parábolas" en el sentido ordinario.)

Para la función $y = e^x$ la Fig. 5.1 muestra las tres primeras parábolas osculadoras en el punto $x = 0$.

Dos curvas, $y = f(x)$ e $y = g(x)$, que tienen contacto de orden n en un punto $x = a$, podrían posiblemente tener contacto de orden aún mayor; esto es, la ecuación $f^{(n+1)}(a) = g^{(n+1)}(a)$ podría también ser cierta. Si no es este el caso, es decir, si $f^{(n+1)}(a) \neq g^{(n+1)}(a)$, se dice que el orden del contacto es exactamente n.[1]

Contacto de orden par o impar

Usando nuestras fórmulas y la intuición, podemos establecer un hecho notable a menudo no observado por los principiantes. Supóngase que el contacto entre dos curvas es exactamente de orden par; esto es, un número par n de derivadas de las dos funciones tienen el mismo valor en el punto en cuestión, en tanto que las $(n + 1)$-ésimas derivadas difieren entre sí. Las fórmulas anteriores muestran entonces que la diferencia $f(a + h) - g(a + h)$ tienen signos diferentes para pequeños valores positivos de h y también para valores negativos numéricamente pequeños de h. Por tanto, las dos curvas *se cruzan* en el punto de contacto. Esto ocurre, por ejemplo, para el contacto de segundo orden si las terceras derivadas tienen valores diferentes. Por el contrario, contactos de orden precisamente impar (por ejemplo, un contacto ordinario de primer orden) implica que la diferencia $f(a + h) - g(a + h)$ tenga el mismo signo para todos los valores de h positivos o negativos numéricamente pequeños; por consiguiente, las dos curvas no se cruzan en una vecindad del punto de contacto. El ejemplo más simple es el contacto de una curva con su tangente. La tangente puede cruzar la curva sólo en puntos en que el contacto es por lo menos de segundo orden; cruzará efectivamente la curva en puntos en que el contacto es par, por ejemplo, un punto de inflexión ordinario, en el cual $f''(x) = 0$ pero $f'''(x) \neq 0$. En los puntos en que el contacto es de orden impar la tangente no cruza la curva; esto ocurre, por ejemplo, en un punto ordinario de la curva en que la segunda derivada no es cero, tal como para la curva $y = x^4$ en el origen.

Del capítulo 4, p. 378, se sabe que para la circunferencia de curvatura en el punto $x = a$, dado por la función $y = g(x)$ en una vecindad de ese punto, no sólo se tiene $g(a) = f(a)$ y $g'(a) = f'(a)$, sino también $g''(a) = f''(a)$. Por tanto, la circunferencia de curvatura es al mismo tiempo la circunferencia osculadora de la curva en el punto considerado, es decir, es la que en ese punto tiene contacto de segundo orden con la curva. En el caso límite de un punto de inflexión o, en general, de un punto en el cual la curvatura sea cero y el radio de curvatura infinito, la circunferencia de curvatura degenera en la tangente. En casos ordi-

---

[1] Que el orden del contacto entre dos curvas es una genuina relación geométrica no afectada por cambios de ejes, es un hecho que puede confirmarse fácilmente por medio de las fórmulas para cambio de ejes (ver capítulo 4, p. 378).

narios, cuando el contacto en el punto que se considera no es de orden superior al segundo, la circunferencia de curvatura no sólo toca la curva sino que también la cruza (ver Fig. 4.23, p. 377).

Como conclusión, se debe mencionar que a veces un contacto de orden exactamente igual a m se describe diciendo: las curvas tienen $m + 1$ puntos en común infinitamente próximos. Por supuesto, el significado preciso de tal afirmación se refiere obviamente a un proceso de límite. Si las curvas tienen en efecto $m + 1$ puntos distintos $P, P_1, \ldots, P_m$ en común y si se deja que todos los puntos P_i tiendan a P, modificando una de las curvas si es necesario, entonces podría esperarse que la posición límite fuese la de dos curvas con un contacto de orden m. Por ejemplo, si se dibuja una circunferencia que pase por tres puntos P, P_1 P_2 de una curva C y luego se hace que P_1 y P_2 tiendan a P, puede verse que la circunferencia tiende a la circunferencia de curvatura de C en P. (Ver problema 4, p. 455.)

b. Sobre la teoría de máximos y mínimos relativos

Como ya se ha visto en el capítulo 3, pp. 264-265, una función $f(x)$ cuya primera derivada se anula en $x = a$ tiene un máximo relativo en ese punto si $f''(a)$ es negativo y un mínimo si $f''(a)$ es positivo. Estas condiciones son, por tanto, suficientes para la existencia de un máximo o un mínimo. Pero de ninguna manera son necesarias, pues en el caso en que $f''(a) = 0$ existen tres posibilidades: en ese punto la función puede tener un máximo, un mínimo o ninguna de las dos cosas. Ejemplos de las tres posibilidades se tienen en las funciones $y = -x^4$, $y = x^4$ e $y = x^3$ en el punto $x = 0$. El teorema de Taylor nos permite de inmediato establecer un teorema general sobre condiciones suficientes para máximo o mínimo. Solamente se necesita desarrollar la función $f(a + h)$ en potencias de h. El punto esencial es entonces determinar si el primer término que no se anula contiene una potencia par o impar de h. En el primer caso se tiene un máximo o un mínimo según que el coeficiente de h sea negativo o positivo; en el segundo caso se tiene una tangente de inflexión horizontal y no se tiene ni máximo ni mínimo. El lector puede por sí mismo completar el argumento usando la fórmula para el residuo.[1]

---

[1] La condición necesaria y suficiente dada antes (p. 263) es, sin embargo, más general y más conveniente para las aplicaciones: suponiendo que la primera derivada $f'(x)$ se anule solamente en un número finito de puntos, una condición necesaria y suficiente para la existencia de un máximo o de un mínimo en esos puntos es que la primera derivada $f'(x)$ cambie de signo conforme la curva pasa a través del punto.

Apéndice I

A.I.1 Ejemplo de una función que no se puede desarrollar en una serie de Taylor

La posibilidad de expresar una función mediante una serie de Taylor con residuo de orden $(n + 1)$ depende esencialmente de la continuidad y la derivabilidad de la función en el punto considerado. Por esta razón $\log x$ no puede representarse por medio de una serie de Taylor en potencias de x; y lo mismo es cierto para la función $x^{1/3}$, cuya derivada es infinita en $x = 0$.

Para que una función pueda ser desarrollada en una serie de Taylor infinita todas sus derivadas deben necesariamente existir en el punto en cuestión. Esta condición, sin embargo, de ninguna manera es suficiente. Una función tal que todas sus derivadas existan y sean continuas en todo un intervalo no necesariamente es desarrollable en serie de Taylor; esto es, el residuo R_n en el teorema de Taylor puede no tender a cero al crecer n, no importa cuán pequeño sea el intervalo en el cual se desee desarrollar la función.

Un ejemplo importante y sencillo de este fenómeno lo constituye la función

$$y = f(x) = e^{-1/x^2} \quad \text{para} \quad x \neq 0, \; f(0) = 0,$$

que ya ha sido considerada en el apéndice del capítulo 3, p. 276. Esta función y sus derivadas son continuas en cualquier intervalo, aún en $x = 0$, como ya se ha visto, y en este punto todas las derivadas se anulan, es decir, $f^{(n)}(0) = 0$ para todo valor de n. (Geométricamente esto significa que la recta $y = 0$ tiene *contacto de orden infinito* con la curva en el punto $x = 0$). Por tanto, en el desarrollo de Taylor

$$f(0) + \frac{x}{1!}f'(0) + \frac{x^2}{2!}f''(0) + \cdots$$

todos los coeficientes del polinomio de aproximación $P_n(x)$ se anulan sin importar qué valor se escoja para n. Así, el residuo permanece igual a la función misma y, por tanto, excepto para $x = 0$, no puede tender a cero conforme n crece, ya que la función es positiva para todo otro valor de x.

De paso se observa que esta función es útil para la construcción de funciones que exhiben fenómenos no esperados por la intuición. Por ejemplo, la función

$$g(x) = e^{-1/x^2} \operatorname{sen}(1/x)$$

complementada con $g(0) = 0$ es una función con derivadas de todo orden, las cuales se anulan en $x = 0$. La gráfica de $y = g(x)$ cerca de $x = 0$ intersecta el eje x una infinidad de veces oscilando otras tantas.

A.I.2 Ceros e infinitos de funciones

a. Ceros de orden n

El desarrollo de Taylor de una función $f(x)$ permite caracterizar el orden con el cual se anula la función en un punto $x = a$. Se dice que una función $f(x)$ tiene un cero de multiplicidad exactamente n en $x = a$, o que se anula ahí exactamente con orden n, si $f(a) = 0, f'(a) = 0, f''(a) = 0, \ldots, f^{(n-1)}(a) = 0$ y $f^{(n)}(a) \neq 0$. Se supone expresamente que en la vecindad del punto la función tiene derivadas continuas al menos hasta de orden n. De la definición se sigue que la serie de Taylor para la función en la vecindad de dicho punto puede escribirse en la forma

$$(28) \qquad f(a+h) = \frac{h^n}{n!}F(h) = \frac{h^n}{n!}f^{(n)}(a+\theta h), \qquad \theta < 1,$$

expresión en la cual, conforme h tiende a cero, el factor $F(h) = n!f(a+h)/h^n$ tiende a un límite diferente de cero, a saber el valor $f^{(n)}(a)$. Por consiguiente, $f(a+h)$ tiene el mismo orden que h^n para $h \to 0$, o se anula con orden n (en el sentido de la definición dada en el capítulo 3, p. 273).

Análogamente, desarrollando las derivadas $f'(x), f''(x), \ldots, f^{(\nu)}(x)$ mediante el teorema de Taylor con la forma de Lagrange para el residuo, se obtiene una serie de expresiones

$$(29) \qquad \begin{aligned} f'(a+h) &= \frac{h^{n-1}}{(n-1)!}F_1(h) = \frac{h^{n-1}}{(n-1)!}f^{(n)}(a+\theta h), \\ f^{\nu}(a+h) &= \frac{h^{n-\nu}}{(n-\nu)!}F_\nu(h) = \frac{h^{n-\nu}}{(n-\nu)!}f^{(n)}(a+\theta h), \end{aligned}$$

en todas las cuales los factores θ *pueden ser diferentes* en tanto que los factores $F_1, F_2, \ldots, F_\nu$ tienden *continuamente* a $f^{(n)}(a)$ conforme $h \to 0$. Por tanto, f' se anula con orden $n-1$, f'' con orden $n-2$, etc.

En estas fórmulas, por supuesto, se hace la hipótesis de que $f(x)$ se anula con orden $n \geqslant \nu$.

b. Infinito de orden ν

Si una función $\phi(x)$ está definida en una vecindad del punto $x = a$, excepto quizá en ese punto, y si puede escribirse $\phi(x) = f(x)/g(x)$ de manera que en $x = a$ el numerador no se anule pero el denominador tenga un cero de multiplicidad ν, se dice que $\phi(x)$ se vuelve infinita de orden ν en el punto $x = a$. Si en $x = a$ el numerador tiene un cero de multiplicidad μ

y si $\mu > \nu$, la función tiene en ese punto un cero de multiplicidad $\mu - \nu$; si $\mu < \nu$, la función tiene una infinidad de orden $\nu - \mu$ en el punto en cuestión.

Estas definiciones están de acuerdo con las convenciones adoptadas antes (ver sección 3.7) relativas al comportamiento de una función.

A.I.3 Expresiones indeterminadas

Discutiremos ahora de manera más precisa las *"expresiones indeterminadas"* de la forma $\phi(x) = f(x)/g(x)$ tal que $f(x)$ y $g(x)$ se anulan ambas en el mismo punto $x = a$, como ocurre para la función $(\operatorname{sen} x)/x$ en $x = 0$. A tales funciones les asignaremos siempre el valor

$$\phi(a) = \lim_{h \to 0} \phi(a + h) \tag{30}$$

suponiendo que este límite exista.

Estos valores límite pueden ser caracterizados mediante una sencilla regla, conocida como *regla de L'Hospital,* para lo cual supondremos la continuidad, en un intervalo que contenga al punto a, de todas las derivadas de f y g que aparezcan. Supondremos además que el denominador $g(x)$ se anula en $x = a$ con orden ν no mayor que el orden correspondiente para el numerador, de manera que la función $\phi(x)$ no se vuelva infinita en $x = a$. La regla establece entonces que

$$\phi(a) = \frac{f^{(\nu)}(a)}{g^{(\nu)}(a)}. \tag{31}$$

Por la definición de continuidad, la función $\phi(x)$ es entonces continua en $x = a$ y, puesto que es continua en todas partes, siempre que $g(x) \neq 0$, es continua en un intervalo en torno de a.

La prueba se sigue inmediatamente de los resultados de A.2. Aplicando las Ecs. (28) tanto a f como a g, se encuentra que la función ϕ está dada. en una vecindad de a, por la relación

$$\phi(a + h) = \frac{f(a + h)}{g(a + h)} = \frac{f^{(\nu)}(a + \theta h)}{g^{(\nu)}(a + \theta_1 h)},$$

de la cual se sigue (31) al considerar la continuidad del numerador y del denominador y el hecho de que $g^{(\nu)}(a)$ no se anula. El significado de las últimas ecuaciones puede expresarse de la siguiente manera: si el numerador y el denominador de una función $\phi(x) = f(x)/g(x)$ se anulan en $x = a$, puede determinarse el valor límite de la función para $x \to a$ derivando numerador y denominador el mismo número de veces hasta que al menos una de las derivadas no se anule en ese punto. Si se encuentra en el denominador una derivada no nula antes que en el numerador, la frac-

ción tiende a cero. Por el contrario, si la derivada no nula aparece primero en el numerador, el valor absoluto de la fracción crece más allá de toda cota.

Se tiene así un método para evaluar la llamada "expresión indeterminada" 0/0, esto es, para determinar el límite de un cociente en el que tanto el numerador como el denominador tienden a cero.

Estos resultados pueden obtenerse de manera un tanto diferente si se basa la deducción en el teorema generalizado del valor medio en vez del teorema de Taylor (ver p. 243). Según esto, si $g'(x) \neq 0$ en una vecindad del punto a, se tiene

$$\frac{f(a+h) - f(a)}{g(a+h) - g(a)} = \frac{f'(a+\theta h)}{g'(a+\theta h)},$$

donde θ es el mismo en numerador y denominador. Por tanto, en el caso particular en que $f(a) = 0 = g(a)$,

$$\frac{f(a+h)}{g(a+h)} = \frac{f'(a+\theta h)}{g'(a+\theta h)}.$$

Aquí θ es un valor en el intervalo $0 < \theta < 1$, y si se hace $k = \theta h$ se obtiene

$$\lim_{h\to 0} \frac{f(a+h)}{g(a+h)} = \lim_{k\to 0} \frac{f'(a+k)}{g'(a+k)},$$

suponiendo que el límite de la derecha existe.

Si $f'(a) = 0 = g'(a)$, se procede de la misma manera hasta llegar a un primer índice μ para el cual ya no se sigue teniendo $f^{(\mu)}(a) = 0 = g^{(\mu)}(a)$. Entonces

$$\lim_{h\to 0} \frac{f(a+h)}{g(a+h)} = \lim_{l\to 0} \frac{f^{(\mu)}(a+l)}{g^{(\mu)}(a+l)} = \frac{f^{(\mu)}(a)}{g^{(\mu)}(a)},$$

expresión que incluye el caso en que ambos miembros son infinitos.

Ejemplos. Los siguientes ejemplos, importantes por sí mismos, ilustran la aplicación de la regla de L'Hospital.

$$\lim_{x\to 0} \frac{\operatorname{sen} x}{x} = \frac{\cos 0}{1} = 1;$$

$$\lim_{x\to 0} \frac{1 - \cos x}{x} = \frac{\operatorname{sen} 0}{1} = 0;$$

$$\lim_{x\to 0} \frac{e^{2x} - 1}{\log(1+x)} = \lim_{x\to 0} \frac{2e^{2x}}{1/(1+x)} = 2;$$

$$\lim_{x\to 0} \frac{1 - \cos x}{x^2} = \lim_{x\to 0} \frac{\operatorname{sen} x}{2x} = \lim_{x\to 0} \frac{\cos x}{2} = \frac{1}{2}.$$

Otras formas indeterminadas. Como siguiente paso se observa que otras de las llamadas formas indeterminadas pueden también reducirse al caso ya considerado; por ejemplo, el límite de

$$\frac{1}{\operatorname{sen} x} - \frac{1}{x}$$

cuando x tiende a cero es el límite de la diferencia de dos expresiones que se hacen infinitas, o bien, es una "forma indeterminada" $\infty - \infty$. Mediante la transformación

$$\frac{1}{\operatorname{sen} x} - \frac{1}{x} = \frac{x - \operatorname{sen} x}{x \operatorname{sen} x}$$

se llega de inmediato a una expresión cuyo límite, cuando x tiende a cero, puede encontrarse aplicando la regla. Así resulta

$$\lim_{x \to 0} \frac{1 - \cos x}{x \cos x + \operatorname{sen} x} = \lim_{x \to 0} \frac{\operatorname{sen} x}{2 \cos x - x \operatorname{sen} x} = 0.$$

Derivadas de formas indeterminadas

Las expresiones $\phi(x) = f(x)/g(x)$ definidas por la regla en $x = a$ no solamente son continuas sino que tienen derivadas también continuas, suponiendo que f y g tienen derivadas continuas de orden suficientemente alto.

Bastará aquí establecer este hecho en el caso en que g se anula en a con primer orden, o sea, $g(a) = 0$, $g'(a) \neq 0$. Para $x \neq a$,

$$\phi'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{(g(x))^2} = \frac{z(x)}{N(x)},$$

donde tanto el numerador como el denominador se anulan en $x = a$, pues $f(a) = g(a) = 0$. Por tanto, el valor límite puede determinarse aplicando la regla

$$\lim_{x \to a} \phi'(x) = \lim_{x \to 0} \frac{z'(x)}{N'(x)}.$$

Claramente, $d(N(x))/dx = 2g(x)g'(x)$, $d(z(x))/dx = g(x)f''(x) - f(x)g''(x)$, las cuales se anulan de nuevo para $x = a$. Aplicando la regla de L'Hospital una vez más,

$$\lim_{x \to a} \phi'(x) = \frac{z''(x)}{N''(x)},$$

y observando que $N''(x) = 2g(x)g''(x) + 2(g'(x))^2$, que no se anula en $x = a$, se encuentra que

$$\lim_{x \to a} \phi'(x) = \frac{g'(a)f''(a) - f'(a)g''(a)}{2(g'(a))^2},$$

y este límite es, sin duda, la derivada de $\phi'(x)$ en $x = a$ (ver capítulo 3, p. 282).

Reglas análogas para formas indefinidas valen para $x \to \infty$. Así, por ejemplo, sean $f(x)$ y $g(x)$ funciones tales que $\lim_{x\to\infty} f(x) = \lim_{x\to\infty} g(x) = 0$, en tanto que $\lim_{x\to\infty} f'(x)$ y $\lim_{x\to\infty} g'(x)$ existen y son diferentes de cero. Entonces

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \frac{\lim_{x\to\infty} f'(x)}{\lim_{x\to\infty} g'(x)}$$

La demostración surge otra vez del teorema del valor medio del cálculo diferencial.

*A.I.4 La convergencia de la serie de Taylor para una función con derivadas no negativas de todos los órdenes

Se incluye a continuación un teorema general sobre la convergencia del desarrollo de Taylor para funciones cuyas derivadas son todas no negativas.

Considérese la clase de funciones $f(x)$ derivables hasta cualquier orden en el intervalo cerrado $a \leq x \leq b$ y tales que todas sus derivadas son no negativas en este intervalo:

$$f^{(\nu)}(x) \geq 0, \qquad \nu = 1, 2, \ldots .$$

Demostraremos que: Para cualquiera de tales funciones el correspondiente desarrollo de Taylor de $f(x + h)$ en potencias de h converge, y la serie representa el valor de $f(x + h)$ cuando x y $\xi = x + h$ están en el intervalo abierto (a, b) y $|h| < b - x$.

La demostración se inicia observando que $f'(x) \geq 0$, por hipótesis, y que, en consecuencia,

$$0 \leq f(x) - f(a) = \int_a^x f'(\xi)\, d\xi$$

$$\leq \int_a^b f'(\xi)\, d\xi = f(b) - f(a) = M.$$

Además, para x y $\xi = x + h$ en el intervalo (a, b), puede escribirse

$$f(x + h) - f(x) = hf'(x) + \cdots + \frac{f^{(n)}(x)}{n!} h^n + R_n.$$

Supóngase primero que $h > 0$, o sea, $x < \xi < b$. Entonces todos los términos del miembro derecho son no negativos,[1] de manera que ninguno de ellos

[1] Esto se deduce para R_n partiendo de las fórmulas de Cauchy o de Lagrange y con la hipótesis $f^{(n+1)} \geq 0$.

es mayor que el valor del miembro izquierdo o que M. Así resulta que

$$0 \leq \frac{f^{(n)}(x)}{n!} \leq \frac{M}{k^n} = \frac{M}{(\xi - x)^n}.$$

Para $\xi \longrightarrow b$ surge que

$$\text{(32)} \qquad \frac{f^{(n)}(x)}{n!} \leq \frac{M}{(b - x)^n}.$$

Ahora bien, por la fórmula de auchy ((19), p. 468) para el residuo, se sabe que existe algún θ en el intervalo $0 \leq \theta \leq 1$ tal que

$$0 \leq R_n = \frac{(1-\theta)^n}{n!} h^{(n+1)} f^{(n+1)}(x + \theta h)$$
$$\leq \frac{h^{n+1}(n+1)(1-\theta)^n M}{(b - x - \theta h)^{n+1}}.$$

Puesto que $\xi = x + h < b$, puede escogerse un número positivo p de manera que

$$0 \leq h \leq \frac{b-x}{1+p}, \qquad \text{o sea,} \qquad b - x - \theta h \geq h(1 + p - \theta).$$

Se tiene entonces

$$0 \leq R_n \leq \frac{Mh^{n+1}(n+1)(1-\theta)^n}{h^{n+1}(1+p-\theta)^{n+1}},$$

o sea,

$$0 \leq R_n \leq \frac{M(n+1)}{(1+p-\theta)}\left(\frac{1-\theta}{1-\theta+p}\right)^n \leq \frac{M(n+1)}{p}\frac{1}{(1+p)^n},$$

ya que

$$\frac{1-\theta}{1-\theta+p} = \frac{1}{1+p/(1-\theta)} \leq \frac{1}{1+p} < 1.$$

Se sabe (capítulo 1, p. 93) que $(n+1)/(1+p)^n$ tiende a cero al crecer n, de manera que cuando $0 \leq h < b - x$ y n crece, R_n tiende a cero; así, la serie de Taylor tiende a la función f para $h \geq 0$.

Para h negativo, el que R_n tienda a cero al crecer n se deduce aplicando la fórmula de Lagrange ((21), p. 468) para R_n:

$$|R_n| = \frac{1}{(n+1)!}|h^{n+1}|\,|f^{(n+1)}(x - \theta\,|h|)|.$$

Ahora bien, $f^{(n+2)}$ es no negativa y, por tanto, $f^{(n+1)}$ es monótona no decreciente. Se sigue entonces de la estimación (32) usada antes que

$$\frac{f^{(n+1)}(x - \theta\,|h|)}{(n+1)!} \leq \frac{f^{(n+1)}(x)}{(n+1)!} \leq \frac{M}{(b-x)^{n+1}}.$$

Por consiguiente,

$$|R_n| \leq \left(\frac{|h|}{b - x}\right)^{n+1} M,$$

y así resulta que R_n tiende a cero conforme n crece, cuando

$$0 < -h < b - x.$$

Entonces, para cualquier punto x tal que $a \leq x < b$, el residuo R_n en la serie de Taylor para $f(x + h)$ en potencias de h tenderá a cero siempre que $|h| < b - x$ y $h > -(x - a)$.

Obsérvese que el resultado anterior es válido aun en el caso en que se suponga la desigualdad $f^{(\nu)}(x) \geq 0$ válida sólo para valores de ν suficientemente grandes (digamos, para $\nu > N$, siendo N algún entero, pudiendo tener $f^{(\nu)}(x)$ cualquier signo para $\nu \leq N$). Para demostrar esto solamente es necesario reemplazar en la demostración anterior la función f por la función

$$g(x) = f(x) + M(x - a + 1)^N,$$

donde M es alguna constante positiva. Entonces, $g^{(\nu)}(x) = f^{(\nu)}(x) \geq 0$ para $\nu > N$, y $g^{(\nu)}(x) = f^{(\nu)}(x) + MN(N - 1)\cdots(N - \nu + 1)(x - a + 1)^{N-\nu} \geq f^{(\nu)}(x) + M$ para $\nu \leq N$. Así, $g^{(\nu)}(x) \geq 0$ para todo ν si M se escoge suficientemente grande. Esto prueba que $g(x)$ puede desarrollarse en potencias de x; y el mismo resultado se deduce entonces para la función f, la cual difiere de g sólo por un polinomio.

El teorema de la serie binomial (p. 475) es una consecuencia inmediata de este resultado: Cámbiese ligeramente la notación y considérese primero la función $\phi(x) = (1 - x)^\alpha$ en vez de $(1 + x)^\alpha$. Las derivadas de ϕ están dadas entonces por

$$\phi^{(\nu)}(x) = (-1)^\nu \binom{\alpha}{\nu} (1 - x)^{\alpha-\nu} \nu!$$

Puesto que los coeficientes binomiales,

$$\binom{\alpha}{\nu} = \frac{\alpha(\alpha - 1)\cdots(\alpha - \nu + 1)}{\nu!},$$

tienen signos alternantes cuando $\alpha - \nu$ es negativo, una de las dos funciones, $\phi(x)$ o $-\phi(x)$, pertenece a la clase de funciones con derivadas no negativas a partir de cierto orden, para x limitada a valores $x < 1$. Así, para $a = -1$, $b = 1$, $x = 0$ y $|h| < b - x = 1$ nuestro teorema general afirma que

$$(1 - h)^\alpha = \sum_{\nu=0}^{\infty} (-1)^\nu \binom{\alpha}{\nu} h^\nu.$$

Si aquí se escribe x en vez de $-h$, se obtiene el desarrollo binomial

$$(1+x)^{\alpha} = \sum_{\nu=0}^{\infty} \binom{\alpha}{\nu} x^{\nu} = 1 + \alpha x + \frac{\alpha(\alpha-1)}{1\cdot 2} x^2 + \frac{\alpha(\alpha-1)(\alpha-2)}{1\cdot 2\cdot 3} x^3 + \cdots$$

para cualquier exponente α y cualquier x tal que $-1 < x < 1$.

Apéndice II Interpolación

*A.II.1 El problema de la interpolación. Unicidad

El polinomio de Taylor $P_n(x)$ aproxima a la función $f(x)$ de manera tal que las gráficas de $f(x)$ y $P_n(x)$ tienen contacto de orden n en un punto a, o sea, de manera que $f(x)$ y $P_n(x)$ coinciden en $n+1$ puntos "infinitamente próximos" a a. Podría "descomponerse" el punto con abscisa a en $n+1$ puntos distintos con abscisas $x_0, x_1, \ldots, x_n$ y buscarse una aproximación para $f(x)$ mediante un polinomio $\phi(x)$ de grado n que coincida con $f(x)$ en estos puntos. Este polinomio, como es fácil ver, está determinado unívocamente por un sistema de ecuaciones lineales. Mediante el paso al límite, $x_i \to x$ para todo i, se obtienen de nuevo los polinomios de Taylor. La "interpolación", esto es, la aproximación mediante polinomios coincidentes con $f(x)$ en puntos distintos, es de gran importancia en muchas aplicaciones. La siguiente discusión proporciona una idea sumaria de la teoría de la interpolación.

Considérese el siguiente problema: Determinar un polinomio $\phi(x)$ de grado n tal que en $n+1$ puntos distintos, $x_0, x_1, \ldots, x_n$, tome los $n+1$ valores dados $f_0, f_1, \ldots, f_n$, o sea,

$$\phi(x_0) = f_0, \quad \phi(x_1) = f_1, \ldots, \quad \phi(x_n) = f_n.$$

Si los números f_i son los valores $f_i = f(x_i)$ que toma una función dada, $f(x)$ (posiblemente menos elemental), en los puntos x_i, entonces el polinomio $\phi(x)$ se llamará el *polinomio de interpolación* de n-ésimo grado de la función $f(x)$ para los puntos $x_0, x_1, \ldots, x_n$.

Existe a lo más solamente uno de tales polinomios de grado n, pues si existiesen dos de ellos, $\phi(x)$ y $\psi(x)$ distintos, su diferencia $D(x) = \phi(x) - \psi(x)$ sería un polinomio de grado m, con $0 \leq m \leq n$ y con $n+1$ raíces distintas, lo cual no es posible según el álgebra elemental.[1]

Puede probarse la *unicidad* del polinomio de interpolación aún por otro método, basado en el

---

[1] Pues se tiene

$$D(x) = c_0(x-x_1)(x-x_2)\ldots(x-x_m), \qquad c_0 \neq 0,$$

ya que $x_1, \ldots, x_m$ son ceros de $D(x)$; por tanto, como $D(x_0) = 0$,

$$c_0(x_0-x_1)(x_0-x_2)\ldots(x_0-x_m) = 0,$$

contrario al hecho de que $x_0, x_1, \ldots, x_m$ son todos distintos.

TEOREMA GENERAL DE ROLLE. *Si una función $F(x)$ tiene derivadas continuas hasta el orden n en un intervalo y se anula por lo menos en $n + 1$ puntos distintos, $x_0, x_1, \ldots, x_n$, del mismo, entonces existe un punto en el interior de dicho intervalo para el cual $F^{(n)}(\xi) = 0$.*

DEMOSTRACIÓN. El teorema general se deduce fácilmente del caso especial en que $n = 1$, que es el teorema de Rolle demostrado en la p. 196. Supónganse los números $x_0, x_1, \ldots, x_n$ dispuestos en orden creciente. Entonces, por el teorema del valor medio (o por el teorema de Rolle), la primera derivada $F'(x)$ debe anularse por lo menos una vez dentro de cada uno de los n subintervalos (x_i, x_{i+1}). Esta misma consideración aplicada a $F'(x)$ y a los intervalos entre sus ceros nos dice que $F''(x)$ se anula en $n - 1$ puntos. Aplicando repetidamente este argumento se demuestra el teorema.

Apliquemos ahora este teorema a la diferencia

$$\begin{aligned} F(x) = D(x) &= \phi(x) - \psi(x) \\ &= d_0x^n + d_1x^{n-1} + \cdots + d_n, \end{aligned}$$

que, por hipótesis, se anula en $n + 1$ puntos. Obtenemos un punto ξ en el cual la n-ésima derivada se anula. $D^{(n)}(\xi) = 0$. Esta es, precisamente, $n!d_0$, de manera que $d_0 = 0$, y la diferencia dada es un polinomio de grado a lo más igual a $n - 1$ y que se anula en $n + 1$ puntos. Aplicando nuevamente el teorema de Rolle se obtiene $d_1 = 0$; etc. O sea $D(x)$ es idénticamente cero, como se afirmó.

Estas consideraciones pueden extenderse al caso en que los x_i no son todos distintos entre sí, sino que, posiblemente r valores x_i coinciden, es decir, $x_0 = x_1 = \cdots = x_{r-1}$. En el problema de la interpolación para estos casos se requiere entonces que $\phi(x)$ y las derivadas $\phi'(x), \ldots, \phi^{(r-1)}(x)$ tomen valores prefijados para $x = x_0$, y, correspondientemente, también para los otros puntos x_ν. El polinomio $D(x)$ es entonces de la forma $c(x - x_0)^r(x - x_r) \cdots$. El teorema general de Rolle y el teorema de unicidad, así como también las demostraciones, permanecen inalterados en este caso.

A.II.2 Construcción de la solución. Fórmula de interpolación de Newton

Construiremos ahora un polinomio de interpolación $\phi(x)$ de grado n, tal que $\phi(x_0) = f_0, \ldots, \phi(x_n) = f_n$. Para construirlo paso a paso, comenzamos con la constante f_0, que es un polinomio $\phi_0(x)$ de orden 0 que toma el valor $A_0 = f_0$ para todo x y en particular para $x = x_0$. Sumamos a éste

un polinomio de primer orden que se anula para $x = x_0$ y es, por consiguiente, de la forma $A_1(x - x_0)$. A continuación determinamos A_1 de manera que la suma tenga para $x = x_1$ el valor correcto f_1. Al polinomio de primer grado que resulta le llamamos $\phi_1(x)$. Sumamos ahora a $\phi_1(x)$ un polinomio de segundo grado que se anula para $x = x_0$ y $x = x_1$, debiendo por consiguiente ser de la forma $A_2(x - x_0)(x - x_1)$, cuya adición no cambia el comportamiento en estos dos puntos. En seguida se determina el factor A_2 de manera que el polinomio de segundo grado resultante, $\phi_2(x)$, tome también el valor fijado, que es f_2 para $x = x_2$ en este caso. Este procedimiento se continúa hasta cubrir todos los puntos, obteniéndose así el polinomio

$$(33) \qquad \phi(x) = \phi_n(x) = A_0 + A_1(x - x_0) + A_2(x - x_0)(x - x_1) + \cdots + A_n(x - x_0)\cdots(x - x_{n-1}).$$

El método para la obtención de los coeficientes A_i en la expresión de ϕ se aclara substituyendo $x = x_0$, $x = x_1$, $\ldots, x = x_n$ en este mismo orden; y así se obtiene el sistema de $n + 1$ ecuaciones

$$(34) \qquad \begin{aligned} f_0 &= A_0 \\ f_1 &= A_0 + A_1(x_1 - x_0) \\ f_2 &= A_0 + A_1(x_2 - x_0) + A_2(x_2 - x_0)(x_2 - x_1) \\ &\cdots\cdots\cdots\cdots \\ f_n &= A_0 + A_1(x_n - x_0) + \cdots + \\ &\qquad + A_n(x_n - x_0)(x_n - x_1)\cdots(x_n - x_{n-1}). \end{aligned}$$

Obviamente, los coeficientes $A_0, A_1, \ldots, A_n$ pueden determinarse en forma sucesiva de manera tal que satisfagan estas ecuaciones, pudiendo construirse así el polinomio de interpolación.

Cuando los valores x_ν son *equidistantes,* o sea $x_\nu = x_{\nu-1} + h$, el resultado puede escribirse explícitamente en una forma más elegante. Las ecuaciones para los A_i se convierten ahora en

$$(35) \qquad \begin{aligned} f_0 &= A_0 \\ f_1 &= A_0 + hA_1 \\ f_2 &= A_0 + 2hA_1 + 2!\,h^2A_2 \\ f_3 &= A_0 + 3hA_1 + 3\cdot 2h^2A_2 + 3!\,h^3A_3 \\ &\cdots\cdots\cdots\cdots \\ f_n &= A_0 + nhA_1 + \cdots + \frac{n!}{(n-i)!}h^iA_i + \cdots + n!\,h^nA_n. \end{aligned}$$

Las soluciones pueden fácilmente ser expresadas como *diferencias sucesivas* de f: Dada cualquier sucesión (finita o infinita) de términos $f_0, f_1, f_2, \ldots$, llamamos a las expresiones

$$\Delta f_0 = f_1 - f_0, \quad \Delta f_1 = f_2 - f_1, \quad \Delta f_2 = f_3 - f_2, \ldots$$

las *primeras diferencias* de las f_k. Aplicando nuevamente el proceso de diferencias a la sucesión de Δf_k, se obtienen las expresiones

$$\Delta^2 f_0 = \Delta f_1 - \Delta f_0, \qquad \Delta^2 f_1 = \Delta f_2 - \Delta f_1, \qquad \Delta^2 f_2 = \Delta f_3 - \Delta f_2, \ldots,$$

o sea,

$$\Delta^2 f_0 = f_2 - 2f_1 + f_0, \qquad \Delta^2 f_1 = f_3 - 2f_2 + f_1, \ldots,$$

que son las *segundas diferencias* de las f_k. La n-ésima diferencia se define recursivamente como $\Delta^{n-1} f_{k+1} - \Delta^{n-1} f_k$. Y cuando se la expresa directamente en términos de las f_k, está dada por la fórmula

$$(36) \qquad \Delta^n f_k = f_{k+n} - \binom{n}{1} f_{k+n-1} + \binom{n}{2} f_{k+n-2} - \cdots + (-1)^n f_k,$$

obtenida mediante un simple argumento inductivo que se deja al lector como ejercicio. Con esta terminología, los coeficientes A_ν pueden escribirse en la forma

$$(37) \qquad A_\nu = \frac{1}{\nu!} h^{-\nu} \Delta^\nu f_0,$$

como puede verificarse por inducción.[1]

Fórmula de interpolación de Newton. Haciendo $\xi = (x - x_0)/h$ se tiene $x - x_r = h(\xi - r)$. Las expresiones $(x - x_0)(x - x_1) \cdots (x - x_\nu)$ adoptan entonces la forma $\xi(\xi - 1) \cdots (\xi - \nu)h^{\nu+1}$. Así se obtiene para los polinomios $\phi(x)$, a partir de (33), (37), la *fórmula de interpolación de Newton:*[2]

$$\phi(x) = \phi(x_0 + \xi h) = f_0 + \binom{\xi}{1} \Delta f_0 + \binom{\xi}{2} \Delta^2 f_0 + \cdots + \binom{\xi}{n} \Delta^n f_0.$$

Si $f_0, f_1, f_2, \ldots$ son los valores de una función $f(x)$ en los puntos $x_0, x_1, x_2, \ldots$, donde f tiene derivadas continuas hasta la de orden n, entonces $\Delta^\nu f_0/h^\nu$ es una aproximación a la derivada $f^{(\nu)}(x_0)$. En la p. 495 se demostrará que

$$\lim_{h\to 0} \frac{1}{h^\nu} \Delta^\nu f_0 = f^{(\nu)}(x_0).$$

---

[1] Debe verificarse que los valores A_ν dados por (37) satisfacen las ecuaciones (35); o sea, que para cualquier sucesión $f_0, f_1, f_2, \ldots$, la identidad

$$f_k = f_0 + \binom{k}{1} \Delta f_0 + \binom{k}{2} \Delta^2 f_0 + \cdots + \binom{k}{k} \Delta^k f_0$$

se satisface. Suponiendo que esto se cumple para un cierto k, debe demostrarse que

$$\begin{aligned} f_{k+1} &= f_1 + \binom{k}{1} \Delta f_1 + \binom{k}{2} \Delta^2 f_1 + \cdots \\ &= (f_0 + \Delta f_0) + \binom{k}{1}(\Delta f_0 + \Delta^2 f_0) + \binom{k}{2}(\Delta^2 f_0 + \Delta^3 f_0) + \cdots \\ &= f_0 + \binom{k+1}{1} \Delta f_0 + \binom{k+1}{2} \Delta^2 f_0 + \cdots \end{aligned}$$

que es la identidad para el caso de $k + 1$.

[2] Como en la p. 476, aquí se definen los coeficientes binomiales $\binom{\xi}{k}$ para un ξ general y enteros de posición k, por medio de $\binom{\xi}{k} = \xi(\xi - 1) \cdots (\xi - k + 1)/k!$

Puesto que también

$$\lim_{h \to 0} h^k \binom{\xi}{k} = \frac{(x - x_0)^k}{k!},$$

se ve que en este caso $\phi(x)$ tiende al polinomio de Taylor $P_n(x)$ cuando h tiende a cero.

Obsérvese que de la misma manera es posible la construcción del polinomio de interpolación en el caso en que, por ejemplo, los primeros r valores, $x_0, \ldots, x_{r-1}$, coinciden y se prefijan los correspondientes valores f_0, $f_0', \ldots, f_0^{(r-1)}$ para $\phi(x_0), \phi'(x_0), \ldots, \phi^{(r-1)}(x_0)$, los cuales coinciden con los valores

$$f(x_0), f'(x_0), \ldots, f^{(r-1)}(x_0),$$

siendo f una función dada. Para $\phi(x)$ se escribe la forma

$$\phi(x) = A_0 + A_1(x - x_0) + A_2(x - x_0)^2 + \cdots + A_r(x - x_0)^r + A_{r+1}(x - x_0)^r(x - x_r) + \cdots;$$

y entonces se determinan los A_ν en forma sucesiva a partir de las ecuaciones

$$f_0 = A_0 \qquad f_0' = A_1 \qquad f_0'' = 2A_2$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$f_0^{(r-1)} = (r-1)!\, A_{r-1}$$

$$f_r = A_0 + A_1(x_r - x_0) + \cdots + A_r(x_r - x_0)^r$$

$$f_{r+1} = A_0 + A_1(x_{r+1} - x_0) + \cdots + A_r(x_{r+1} - x_0)^r + A_{r+1}(x_{r+1} - x_0)^r(x_{r+1} - x_r)$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

A.II.3 La estimación del residuo

En vista de las consideraciones anteriores, no importa cómo estén dados originalmente los valores $f_0, f_1, \ldots, f_n$. Por ejemplo, si estos valores fuesen obtenidos mediante observaciones físicas, el problema de construir el polinomio de interpolación estaría resuelto completamente, resultando $\phi(x)$ como una función "suave" o "lisa" definida para todo x, que toma los valores observados para los puntos en cuestión, los cuales podrían usarse para "predecir" valores aproximados de $f(x)$ en otros x. Sin embargo, si la función $f(x)$, que toma en los puntos x_k dados los $n + 1$ valores f_k, está definida también para valores intermedios x, se tiene que encarar el problema de estimar la *diferencia* $R(x) = f(x) - \phi(x)$, llamada *error de interpolación*. Al principio sólo se sabe que $R(x_0) = R(x_1) = \cdots R(x_n) = 0$. Para disponer de mayor información deberán hacerse hipótesis adicionales relativas al comportamiento de la función $f(x)$ y que afecten al residuo $R(x)$. Se

supondrá, por consiguiente, que en el intervalo en consideración $f(x)$ tiene derivadas continuas, al menos hasta la de orden $n + 1$.

Se observa en primer lugar que para cualquier elección de la constante c, la función

$$K(x) = R(x) - c(x - x_0)(x - x_1) \cdots (x - x_n)$$

se anula en los $n + 1$ puntos $x_0, \ldots, x_n$. Escójase ahora cualquier valor y distinto de $x_0, x_1, \ldots, x_n$. Se puede entonces determinar c de manera que $K(y) = 0$, o sea,

$$c = \frac{R(y)}{(y - x_0)(y - x_1) \cdots (y - x_n)}.$$

Existen, por tanto, $n + 2$ puntos en los que $K(x)$ se anula. Aplicando a $K(x)$ el teorema de Rolle generalizado, se ve que existe un valor $x = \xi$ entre el mayor y el menor de los valores $x_0, x_1, \ldots, x_n$, y, tal que $K^{(n+1)}(\xi) = 0$. Puesto que $R(x) = f(x) - \phi(x)$, y en vista de que $\phi(x)$, por ser un polinomio de orden n, tiene derivada $(n + 1)$-ésima idénticamente nula, resulta que

$$f^{(n+1)}(\xi) - c(n + 1)! = 0,$$

[habiéndose usado el hecho de que $(n + 1)!$ es la $(n + 1)$-ésima derivada de $(x - x_0) \cdots (x - x_n)$]. Así se ha obtenido para c una segunda expresión, $c = f^{(n+1)}(\xi)/(n + 1)!$, que contiene a ξ y que de alguna manera depende de y. Usando ahora la ecuación $K(y) = 0$, en la cual y es completamente arbitraria y puede entonces ser reemplazada por x, se obtiene la representación

$$(38) \qquad R(x) = \frac{(x - x_0)(x - x_1) \cdots (x - x_n)}{(n + 1)!} f^{(n+1)}(\xi),$$

donde ξ es algún valor comprendido entre el menor y el mayor de los puntos $x, x_0, x_1, \ldots, x_n$.

En esta forma el problema general de la interpolación para una función $f(x)$ dada queda resuelto completamente. Se tiene para $f(x)$ la representación

$$(39) \qquad f(x) = A_0 + A_1(x - x_0) + A_2(x - x_0)(x - x_1) + \cdots + A_n(x - x_0)(x - x_1) \cdots (x - x_{n-1}) + R_n,$$

donde los coeficientes $A_0, A_1, \ldots, A_n$ pueden ser encontrados sucesivamente a partir de los valores de f en los puntos $x_0, x_1, \ldots, x_n$, usando las fórmulas de recurrencia (34) de la p. 491, y donde el residuo R_n es de la forma

$$(40) \qquad R_n = \frac{(x - x_0)(x - x_1) \cdots (x - x_n)}{(n + 1)!} f^{(n+1)}(\xi),$$

siendo ξ un número escogido adecuadamente entre el mayor y el menor de los valores $x, x_0, x_1, \ldots, x_n$.

Si la fórmula (39) que da $f(x)$ se escribe para $n-1$ en vez de n y el resultado se resta de la original, se obtiene

$$A_n(x-x_0)(x-x_1)\cdots(x-x_{n-1})+R_n-R_{n-1}=0.$$

Para $x=x_n$ resulta $R_n=0$ y, por tanto, el coeficiente A_n (usando (40) con $n-1$ en vez de n) tiene la representación

$$A_n=\frac{f^{(n)}(\xi)}{n!}$$

donde ξ es entre el menor y el mayor de los valores $x_0, x_1, \ldots, x_n$. Representaciones análogas existen para $A_{n-1}, A_{n-2}, \ldots, A_0$. En esta forma se ve que si los $x_0, x_1, \ldots, x_n$ se aproximan entre sí tendiendo hacia uno y el mismo punto, posiblemente el origen, la fórmula de interpolación (39) se convierte, término a térmnio, en la fórmula de Taylor ((27*a*), p. 471), con la forma de Lagrande ((21), p. 468) para el residuo. La fórmula de Taylor puede en esta forma considerarse como un caso límite de la fórmula de interpolación de Newton.

Esta fórmula nos permite dar un significado preciso a una expresión usada muy a menudo en geometría. La parábola osculante, que tiene contacto de orden n en un punto con una curva dada, se dice también que posee "$(n+1)$ puntos consecutivos en común" con dicha curva en ese punto. De hecho, esta parábola osculante se obtiene encontrando una parábola que tenga $n+1$ puntos en común con la curva y haciendo después que esos puntos se aproximen entre sí. En forma analítica esto corresponde justamente a la transición del polinomio de interpolación al polinomio de Taylor. De manera semejante puede caracterizarse la condición de osculación de dos curvas arbitrarias. Por ejemplo, la circunferencia de curvatura es aquélla que tiene tres puntos consecutivos en común con la curva dada.

Es de esperarse que la fórmula de interpolación aplicada al caso de una función cuyos valores en puntos definidos se conocen dé los valores de ésta con un alto grado de precisión *entre* dichos puntos (tanto $|f^{(\nu+1)}(\xi)|$ como $|x-x_i|$ son entonces acotadas). Si el valor x cae *fuera* de los intervalos de los puntos $x_0, x_1, \ldots, x_n$, se habla de *extrapolación.* Mediante tal extrapolación puede obtenerse una precisión bastante buena siempre que el punto x esté lo suficientemente próximo a los puntos dados. La fórmula de Taylor, en cierto sentido, corresponde a *extrapolación completa.* En general, sólo es adecuada en la vecindad de un punto.

A.II.4 La fórmula de interpolación de Lagrange

Para terminar, se resuelve el problema de la interpolación usando una fórmula, debida a Lagrange, ligeramente distinta y que difiere de la fórmula de interpolación de Newton por contener cada término individual sólo

uno de los valores dados de la función. Además, la fórmula da $\phi(x)$ en forma completamente explícita, sin necesidad de fórmulas de recurrencia. Se introduce, por brevedad, el polinomio de grado $(n + 1)$,

$$\psi(x) = (x - x_0)(x - x_1)\cdots(x - x_n),$$

que corresponde a los puntos x_ν dados. Derivando con la regla del producto y substituyendo luego sucesivamente x por los valores $x_0, \ldots, x_n$, se obtienen las relaciones

$$\psi'(x_0) = (x_0 - x_1)(x_0 - x_2)\cdots(x_0 - x_n)$$

$$\psi'(x_\nu) = (x_\nu - x_0)\cdots(x_\nu - x_{\nu-1})(x_\nu - x_{\nu+1}),\cdots(x_\nu - x_n),$$

$$\psi'(x_n) = (x_n - x_0)(x_n - x_1)\cdots(x_n - x_{n-1}).$$

Obsérvese que

$$\frac{\psi(x)}{(x - x_\nu)\psi'(x_\nu)} = \frac{(x - x_0)\cdots(x - x_{\nu-1})(x - x_{\nu+1})\cdots(x - x_n)}{(x_\nu - x_0)\cdots(x_\nu - x_{\nu-1})(x_\nu - x_{\nu+1})\cdots(x_\nu - x_n)}$$

es un polinomio de grado n cuyo valor en el punto $x = x$ es 1 y en los puntos x_i restantes es 0. Se ve entonces de inmediato que la expresión

$$(41) \qquad \phi(x) = \psi(x)$$
$$\left[\frac{f_0}{(x - x_0)\psi'(x_0)} + \frac{f_1}{(x - x_1)\psi'(x_1)} + \cdots + \frac{f_n}{(x - x_n)\psi'(x_n)}\right]$$

es el polinomio de interpolación deseado. Esta es la *fórmula de interpolación de Lagrange.*

PROBLEMAS

SECCION 5.4*b*, página 469

1. Dar la derivación formal completa de la fórmula del residuo (27), de p. 471, usando inducción matemática.

2. *(Una variante de la demostración del teorema de Taylor)*

(a) Si $g(h)$ tiene sus primeras $(n + 1)$ derivadas continuas en el intervalo $0 \leq h \leq A$ y si $g(0) = g'(0) = \cdots = g^{(n)}(0) = 0$, en tanto que $|g^{(n+1)}(h)| \leq M$ en $[0, A]$, donde M es una constante, demostrar que $|g^{(n)}(h)| \leq Mh$, $|g^{(n+1)}(h)| \leq Mh^2/2!, \ldots, |g^{(n-i)}(h)| \leq Mh^i/i!, \ldots, |g(h)| \leq Mh^n/n!$, para todo h en el intervalo.

(b) Sean $f(x)$ una función suficientemente derivable en $a \leq x \leq b$, y $T_n(h)$ el polinomio de Taylor para $f(x)$ en $x = a$. Aplíquese el resultado de (a) a la función $g(h) = R_n = f(a + h) - T_n(h)$ para obtener la fórmula de Taylor con una estimación grosera para el residuo.

3. Supóngase que $f(x)$ tiene una derivada continua en el intervalo $a \leq x \leq b$, y sea $f''(x) \geq 0$ para todo valor de x. Entonces, si ξ es cualquier punto en el inter-

valo, la curva en ninguna parte queda por debajo de su tangente en el punto $x = \xi$, $y = f(\xi)$.

(Usar el desarrollo de Taylor hasta el tercer término.)

4. Deducir la fórmula integral para el residuo R_n integrando por partes

$$f(x + h) - f(x) = \int_0^h f'(x + \tau)\, d\tau.$$

5. Integrar por partes la fórmula

$$R_n = \frac{1}{n!}\int_0^h (h - \tau)^n f^{(n+1)}(x + \tau)\, d\tau,$$

y así obtener

$$R_n = f(x + h) - f(x) - hf'(x) - \cdots - \frac{h^n}{n!} f^{(n)}(x).$$

*6. Supóngase que de alguna manera se ha obtenido una serie para la función $f(x)$, a saber

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n + R_n(x),$$

donde $a_0, a_1, \ldots, a_n$ son constantes, $R_n(x)$ tiene n derivadas continuas y $R_n(x)/x^n \to 0$ cuando $x \to 0$. Demostrar que $a_k = (f^k(0)/k!)$ $(k = 0, \ldots, n)$, esto es, que la serie es una serie de Taylor.

SECCION 5.5, página 472

1. Encontrar los primeros cuatro términos no nulos en la serie de Taylor para las siguientes funciones en la vecindad de $x = 0$:

(a) $x \cot x$ (d) $e^{\operatorname{sen} x}$

(b) $\dfrac{\sqrt{\operatorname{sen} x}}{\sqrt{x}}$ (e) e^{e^x}

(c) $\sec x$ (f) $\log \operatorname{sen} x - \log x$.

2. Encontrar la serie de Taylor para arc sen x en la vecindad de $x = 0$,, usando

$$\operatorname{arc\,sen} x = \int_0^x \frac{dt}{\sqrt{1 - t^2}}.$$

Comparar con el problema 2, sección 3.2.

*3. Encontrar los tres primeros términos no nulos de la serie de Taylor para $\operatorname{sen}^2 x$ en la vecindad de $x = 0$, mediante la multiplicación de la serie de Taylor para $\operatorname{sen} x$ por sí misma. Justificar este procedimiento.

*4. Encontrar los tres primeros términos no nulos de la serie de Taylor para $\tan x$ en la vecindad de $x = 0$, usando la relación $\tan x = \operatorname{sen} x/\cos x$. Justificar el procedimiento.

*5. Encontrar los tres primeros términos no nulos de la serie de Taylor para $\sqrt{\cos x}$ en la vecindad de $x = 0$, aplicando el teorema del binomio a la serie de Taylor para $\cos x$. Justificar el procedimiento.

*6. Encontrar la serie de Taylor para $(\operatorname{arc\,sen} x)^2$. Comparar con el problema 2, sección 3.2.

*7. Encontrar las series de Taylor para las siguientes funciones en la vecindad de $x = 0$:

(a) $\operatorname{senh}^{-1} x$. (b) $\displaystyle\int_0^x e^{-t^2}\, dt$. (c) $\displaystyle\int_0^x \frac{\operatorname{sen} t}{t}\, dt$.

***8.** Estimar el error que se comete al usar sólo los n primeros términos en el problema 7.

***9.** La función elíptica $s(u)$ ha sido definida (sección 3.14*a*) como la inversa de la integral elíptica

$$u(s) = \int_0^s \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}.$$

Encontrar el desarrollo de Taylor para $s(u)$ hasta el término de grado 5.

***10.** Evaluar los siguientes límites:

(*a*) $\lim_{x\to\infty} x\left[\left(1+\frac{1}{x}\right)^x - e\right]$,

(*b*) $\lim_{x\to\infty} \left\{\frac{e}{2}x + x^2\left[\left(1+\frac{1}{x}\right)^x - e\right]\right\}$,

*(*c*) $\lim_{x\to\infty} x\left[\left(1+\frac{1}{x}\right)^x - e\log\left(1+\frac{1}{x}\right)^x\right]$.

(*d*) $\lim_{x\to 0} \left(\frac{\operatorname{sen} x}{x}\right)^{1/x^2}$,

(*e*) $\lim_{x\to\infty} \left(\frac{\operatorname{sen} x}{x}\right)^{1/x^2}$

***11.** Encontrar los tres primeros términos de la serie de Taylor para $[1 + (1/x)]^x$ en potencias de $1/x$.

***12.** Dos partículas con cargas de signo opuesto, $+e$, $-e$, situadas a una pequeña distancia d una de la otra, forman un *dipolo* eléctrico con momento $M = ed$. Demostrar que la energía potencial:

(*a*) En un punto situado sobre el eje del dipolo a una distancia r del centro del mismo, es $(M/r^2)(1 + \varepsilon)$, donde ε es aproximadamente igual a $d^2/4r^2$.

(*b*) En un punto situado sobre la bisectriz perpendicular al dipolo, es 0.

(*c*) En un punto con coordenadas polares r, θ, relativas al centro y al eje del dipolo, es $[M\cos(\theta/r^2)](1 + \varepsilon)$, donde ε es aproximadamente igual a

$$(d^2/8r^2)(5\cos^2\theta - 3).$$

(La energía potencial de una carga puntual única q en un punto situado a una distancia r de ella es q/r; y la energía potencial de varias cargas es la suma de las energías potenciales de las cargas por separado.)

SECCION 5.6, página 476

1. Demostrar que si $f(a) = 0$ y $f(x)$ tiene derivadas en $x = a$ hasta de un orden suficientemente alto, entonces $f(x)^n$ tiene al menos un contacto de orden $(n - 1)$ con el eje x.

2. La curva $y = f(x)$ pasa por el origen O y toda al eje x en ese punto. Demostrar que el radio de curvatura de la curva en O está dado por

$$\rho = \lim_{x\to 0} \frac{x^2}{2y}.$$

***3.** K es una circunferencia que toca en un punto P a una curva dada y pasa por un punto cercano Q situado sobre la curva. Demostrar que el límite de circunferencia K para $Q \to P$ es la circunferencia de curvatura de la curva en P.

***4.** Demostrar que el orden de contacto entre una curva y su circunferencia osculante es al menos igual a tres en puntos en que el radio de curvatura es un máximo o un mínimo.

***5.** Demostrar que la circunferencia osculante en un punto en que el radio de curvatura es un máximo o un mínimo no cruza a la curva a menos que el contacto sea de orden superior al tercero.

*6. Encontrar los máximos y mínimos de las siguientes funciones:

(a) $\cos x \cosh x$ $\quad$ (b) $x + \cos x$

*7. Determinar los máximos y mínimos de la función $y = e^{-1/x^2}$ (ver p. 263).

SECCION A.I.3, página 483

1. Demostrar que si f es continua en el intervalo [0, 1], entonces

$$\lim_{x \to 0} x \int_x^1 \frac{f(z)}{z^2}\, dz = f(0).$$

2. Demostrar que la función $y = (x^2)^x$, $y(0) = 1$ es continua en $x = 0$.

6

Métodos numéricos

La tarea de resolver un problema analítico queda siempre inconclusa. Demostrar la existencia y algunas de las propiedades básicas de la solución se considera por lo general satisfactorio, pero quedan siempre importantes preguntas por responder. Así, cuando la solución es definida mediante un proceso de límite, usando una integral, por ejemplo, surge el problema de encontrar realmente aproximaciones numéricas a este límite y de estimar la exactitud de estas aproximaciones. Tales cuestiones no sólo son de importancia básica desde el punto de vista teórico, sino que son también inevitables cuando se desea aplicar el análisis a la descripción y al control de los fenómenos naturales, los cuales pueden describirse en principio sólo de manera aproximada.

Se plantea así el difícil problema de llevar la solución al punto en que puedan obtenerse respuestas numéricas y estimaciones de su exactitud.

Recientemente, con el advenimiento de las máquinas calculadoras automáticas de alta velocidad, los aspectos teóricos y prácticos del "análisis numérico" han recibido un gran impulso. Estos aspectos son presentados en una gran variedad de libros de texto.[1] Durante siglos, sin embargo, muchos de los matemáticos más notables, como Newton, Euler y, en particular, Gauss, han aportado grandes contribuciones a los métodos numéricos.

En este volumen no es posible presentar una visión completa del análisis numérico, pero, al menos, se discutirán algunos de los resultados clásicos más simples.

[1] Véanse, por ejemplo: Hildebrand, *Introduction to Numerical Analysis*, McGraw-Hill Book Co., 1956; Householder, *Principles of Numerical Analysis*, McGraw-Hill Book Co., 1953; Whittaker and Robinson, *The Calculus of Observations*, Blackie and Sons, Ltd., 1929.

6.1 Cálculo de integrales

Aunque la existencia de la integral de una función (continua) es asegurada por la teoría que se dio en el capítulo 2, la evaluación o "cuadratura" [1] de dicha integral no puede realizarse mediante funciones elementales, salvo en casos relativamente raros. Deberán desarrollarse, por consiguiente, métodos de integración numérica y también de estimación de la exactitud de tales aproximaciones numéricas.

Para calcular en forma aproximada la integral

$$J = \int_a^b f(x)\,dx \tag{1}$$

con $a < b$, se subdivide el intervalo $a \leq x \leq b$ en n partes iguales, cada una de longitud $h = (b - a)/n$, mediante los $n + 1$ puntos

$$x_\nu = a + \nu h, \qquad nh = b - a, \qquad \nu = 0, 1, \ldots, n. \tag{2}$$

Entonces,

$$J = \sum_{\nu=1}^{n} J_\nu,$$

donde

$$J_\nu = \int_{x_{\nu-1}}^{x_\nu} f(x)\,dx. \tag{3}$$

El problema de calcular la integral J queda reducido al de obtener buenas aproximaciones para las áreas J_ν de anchura h en las cuales se ha dividido el área total representada por J.

a. Aproximación mediante rectángulos

La aproximación más directa, sugerida por la definición original de la integral, conduce a la relación

$$J = \sum_{\nu=1}^{n} J_\nu,$$

$$\approx h(f_1 + f_2 + \cdots + f_n),$$

donde, por brevedad, se ha hecho

$$f_\nu = f(x_\nu).$$

Aquí (y en todo este capítulo) el símbolo $\approx$ significa "es aproximadamente igual a".

[1] La palabra "cuadratura" indica el proceso de "cuadrar", esto es, de medir el área dentro de una curva encontrando un cuadrado de área igual (como en el problema de "cuadrar el círculo").

Para estimar la exactitud o el "error" de esta aproximación, supóngase que $f(x)$ es continua, con una derivada acotada uniformemente en el intervalo $a \leq x \leq b$: $|f'(x)| \leq M_1$. Puede entonces demostrarse fácilmente (ver problema 4, p. 527, 6.1), que

$$(4) \qquad |J_\nu - hf_\nu| \leq \frac{M_1h^2}{2},$$

o sea,

$$\left| J - h \sum_{\nu=1}^{n} f_\nu \right| \leq n \frac{M_1h^2}{2}$$

$$(5) \qquad = \tfrac{1}{2}M_1(b-a)h.$$

Así, el error de la aproximación para la integral mediante la suma finita es del orden de h, la "anchura de malla", en la terminología del capítulo 3, p. 273.

b. Aproximaciones refinadas — Regla de Simpson

Una mejor aproximación se obtiene casi sin esfuerzo adicional si las áreas J_ν se aproximan, no mediante franjas rectangulares, sino mediante trapezoides angostos, como en la Fig. 6.1*a*. La fórmula de aproximación (fórmula del trapezoide) es entonces

$$J \approx \tfrac{1}{2}h(f_0 + f_1) + \tfrac{1}{2}h(f_1 + f_2) + \cdots + \tfrac{1}{2}h(f_{n-1} + f_n)$$

$$(6) \qquad = h(f_1 + f_2 + \cdots + f_{n-1}) + \frac{h}{2}(f_0 + f_n),$$

pues cada valor de la función, excepto el primero y el último, aparece dos veces.

Una aproximación que resulta por lo general un poco más precisa que la fórmula del trapezoide es aquélla en que la ν-ésima franja se aproxima mediante un tropezoide limitado superiormente por la tangente a la curva en el punto medio $x_{\nu-1} + h/2$ del intervalo $x_{\nu-1} \leq x \leq x_\nu$. El área de este trapezoide es simplemente

$$hf_{\nu-1/2} = hf\left(x_{\nu-1} + \frac{h}{2}\right),$$

y por adición se obtiene la *fórmula de la tangente*,

$$(7) \qquad J \approx h(f_{1/2} + f_{3/2} + \cdots + (f_{(2n-1)/2}).$$

Como se verá en la p. 506, el error de esta aproximación es del orden de h^2 cuando la segunda derivada de f es continua en el intervalo $a \leq x \leq b$ y $|f''(x)| < M_2$, siendo M_2 alguna cota constante.

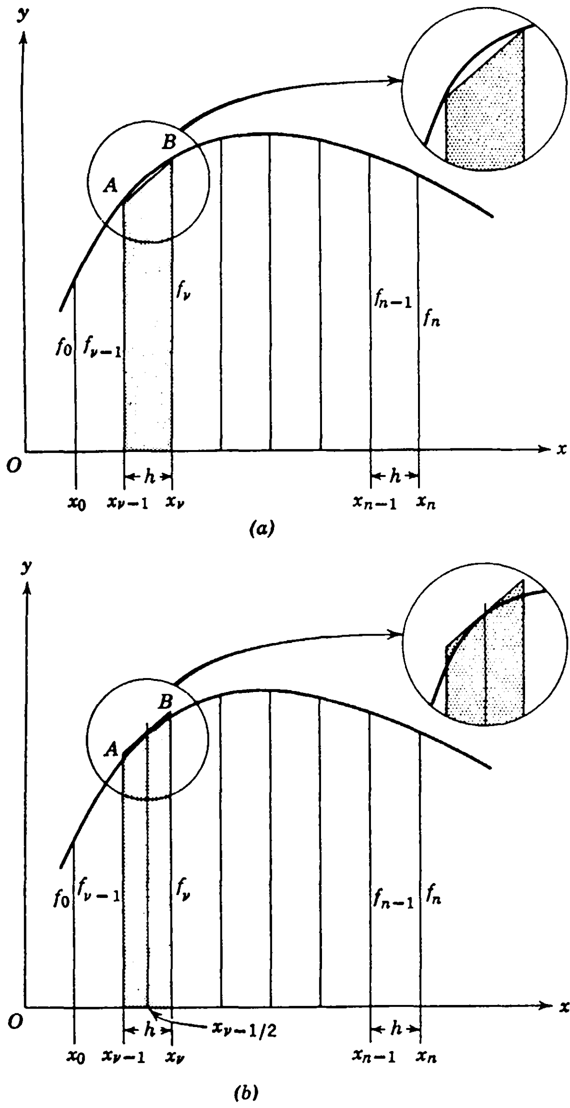


Figura 6.1 (*a*) La fórmula del trapezoide. (*b*) La fórmula de la tangente.

Finalmente, se menciona la famosa aproximación de *Simpson,* la cual con poco mayor trabajo conduce a una aproximación mucho más precisa si la cuarta derivada de f existe y es uniformemente acotada en el intervalo dado.

$$|f^{(4)}(x)| \leq M_4,$$

donde M_4 es una constante. La fórmula de Simpson para $n = 2m$ es

$$(8) \qquad J \approx \frac{4h}{3}(f_1 + f_3 + f_5 + \cdots + f_{2m-1}) + \frac{2h}{3}(f_2 + f_4 + f_6 + \cdots + f_{2m-2}) + \frac{h}{3}(f_0 + f_{2m}).$$

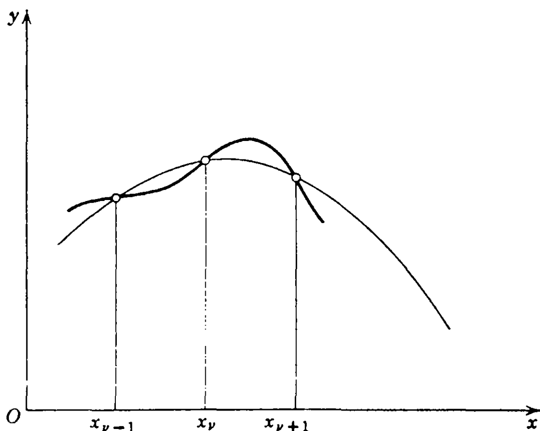


Figura 6.2 Regla de Simpson.

La fórmula es obtenida fácilmente si la región formada por las franjas ν y $(\nu + 1)$ se aproxima mediante una franja de anchura $2h$ limitada superiormente por la parábola que coincide con f para las tres abscisas $x_{\nu-1}$, $x_\nu = x_{\nu-1} + h$ y $x_{\nu+1} = x_{\nu-1} + 2h$ (ver Fig. 6.2). La fórmula de interpolación de Newton (p. 492) conduce a la ecuación de esta parábola:

$$y = f_{\nu-1} + (x - x_{\nu-1}) \cdot \frac{f_\nu - f_{\nu-1}}{h} + \frac{(x - x_{\nu-1})(x - x_{\nu-1} - h)}{2} \frac{f_{\nu-1} - 2f_\nu + f_{\nu-1}}{h^2}.$$

teniéndose, por tanto, la aproximación

$$J_\nu + J_{\nu+1} \approx \int_{x_{\nu-1}}^{x_{\nu+1}} y\,dx = \int_{x_{\nu-1}}^{x\quad +2h} y\,dx$$

$$= 2hf_{\nu-1} + 2h(f_\nu - f_{\nu-1}) + \frac{8/3h - 2h}{2}(f_{\nu+1} - 2f_\nu + f_{\nu-1})$$

$$= \frac{h}{3}(f_{\nu-1} + 4f_\nu + f_{\nu+1}).$$

La fórmula se obtiene ahora para n par $(n = 2m)$ sumando todos estos valores aproximados para $\nu = 1, 3, 4, \ldots, 2m - 1$, o sea, todas las áreas de los pares de franjas.

***Exactitud**

No es difícil estimar la exactitud de las anteriores aproximaciones. Toda cuadratura consiste en aproximar la función $f(x)$ en un intervalo por medio de una función $\phi(x)$ fácilmente integrable (un polinomio). Entonces, una estimación del error de la fórmula de integración puede obtenerse estimando $|f(x) - \phi(x)|$.

En la fórmula de la tangente (p. 503) se reemplaza $f(x)$ en el intervalo $[x_{\nu-1}, x_\nu]$ por su tangente en el punto medio $x_\nu - (h/2)$, esto es, por

$$\phi(x) = f\left(x_\nu - \frac{h}{2}\right) + \left(x - x_\nu + \frac{h}{2}\right) f'\left(x_\nu - \frac{h}{2}\right).$$

Usando el teorema de Taylor con la forma de Lagrange para el residuo:

$$f(x) = \phi(x) + \frac{1}{2}\left(x - x_\nu + \frac{h}{2}\right)^2 f''(\xi),$$

donde ξ está entre x y $x_\nu - h/2$. Por tanto, el error correspondiente a una franja se estima en

$$\begin{aligned} |J_\nu - hf_{\nu-1/2}| &= \left| \int_{x_{\nu-1}}^{x_\nu} [f(x) - \phi(x)]\, dx \right| \\ &\leq \int_{x_{\nu-1}}^{x_\nu} |f(x) - \phi(x)|\, dx \leq M_2 \int_{x_\nu - h}^{x_\nu} \frac{1}{2}\left(x - x_\nu + \frac{h}{2}\right)^2 dx \\ &= \frac{h^3}{24} M_2. \end{aligned}$$

Para el error total en la fórmula de la tangente al considerar todos los intervalos,[1] se encuentra entonces la cota superior

$$n\frac{h^3}{24}M_2 = \frac{h^2}{24}M_2(b - a).$$

Esta derivación se usará como modelo para estimar el error en las otras fórmulas de cuadraturas. Así, en la regla del trapezoide (6) $f(x)$ se aproxima en el intervalo $[x_{\nu-1}, x_\nu]$ por el polinomio de interpolación lineal

$$\phi(x) = f_{\nu-1} + (x - x_{\nu-1})\, \frac{f_\nu - f_{\nu-1}}{h}.$$

Usando la estimación para el error del residuo en la fórmula de interpolación [ver p. 494, Ec. (40)] con $n = 1$, se encuentra

$$f(x) - \phi(x) = \tfrac{1}{2}(x - x_{\nu-1})(x - x_\nu) f''(\xi),$$

[1] Este es el error total inherente al uso de la fórmula de aproximación, el así llamado error "por truncamiento". En la práctica aparecen otros errores debido al redondeo que se hace en los cálculos. El efecto total de los errores por redondeo aumenta generalmente con el número de pasos efectuados, o sea, al disminuir h, en tanto que el error por truncamiento disminuye.

donde ξ está entre $x_{\nu-1}$ y x_ν. Por tanto, el valor absoluto del error en el cálculo de J_ν es a lo más

$$M_2 \int_{x_{\nu-1}}^{x_\nu} |\tfrac{1}{2}(x - x_{\nu-1})(x - x_\nu)|\, dx = \frac{h^3}{12} M_2,$$

y el error total es a lo más n veces la cantidad anterior:

$$\frac{h^2}{12} M_2(b - a).$$

La misma técnica puede aplicarse a la regla de Simpson, (8), tomando como $\phi(x)$ el polinomio cuadrático que coincide con f en los puntos $x_{\nu-1}$, x_ν, $x_{\nu+1}$. Así resulta un error en $J_\nu + J_{\nu+1}$ del orden de h^4. Realmente, sin embargo, puede mejorarse en un orden de magnitud la estimación del error usando un polinomio cúbico $\phi(x)$, el cual da una mejor aproximación que el cuadrático para f en el intervalo $[x_{\nu-1}, x_{\nu+1}]$ y tiene aún la misma integral, conduciendo, por tanto, a la misma fórmula de aproximación (9) para J. Simplemente se usa el polinomio de interpolación que coincide con $f(x)$ en los puntos $x_{\nu-1}$, x_ν, $x_{\nu+1}$ y para el cual $\phi'(x_\nu) = f'(x_\nu)$. Tiene la forma

$$\phi(x) = A_0 + A_1(x - x_{\nu-1}) + A_2(x - x_{\nu-1})(x - x_\nu) + A_3(x - x_{\nu-1})(x - x_\nu)(x - x_{\nu+1}).$$

Aquí los tres primeros términos representan el polinomio de interpolación cuadrático que coincide con f en los tres puntos $x_{\nu-1}$, x_ν, $x_{\nu+1}$. La constante A_3 tiene que determinarse a partir de la condición $\phi'(x_\nu) = f'(x_\nu)$.

El último término,

$$A_3(x - x_\nu + h)(x - x_\nu)(x - x_\nu - h) = A_3[(x - x_\nu)^2 - h^2]\cdot[x - x_\nu],$$

es obviamente una función impar de $x - x_\nu$, y, por consiguiente, no contribuye a la integral entre los límites $x_\nu - h$ y $x_\nu + h$. Para el error en la aproximación de f se tiene entonces la estimación siguiente [ver (40), p. 494, con $n = 3$ y con dos de los puntos de interpolación coincidiendo en x_ν]:

$$f - \phi = \frac{1}{4!}(x - x_{\nu-1})(x - x_\nu)^2(x - x_{\nu+1}) f^{(4)}(\xi).$$

Esto conduce a la siguiente estimación para el error total en el cálculo de $J_\nu + J_{\nu+1}$:

$$\frac{h^5}{90} M_4,$$

y, por tanto, para el error total se tiene la estimación

$$\frac{n}{2}\,\frac{h^5}{90} M_4 = \frac{h^4}{180}(b - a)\, M_4.$$

Naturalmente, puede lograrse una mayor precisión aproximando la función $f(x)$ en una franja por un polinomio de orden aún más alto.

Ejemplos. Apliquemos estos métodos al cálculo de

$$\log_e 2 = \int_1^2 \frac{dx}{x}.$$

Dividiendo el intervalo $1 \leq x \leq 2$ en diez partes de longitud $h = 1/10$ y usando la regla del trapezoide (6) se obtiene

$$\begin{array}{ll}
x_1 = 1.1 & f_1 = 0.90909 \\
x_2 = 1.2 & f_2 = 0.83333 \\
x_3 = 1.3 & f_3 = 0.76923 \\
x_4 = 1.4 & f_4 = 0.71429 \\
x_5 = 1.5 & f_5 = 0.66667 \\
x_6 = 1.6 & f_6 = 0.62500 \\
x_7 = 1.7 & f_7 = 0.58824 \\
x_8 = 1.8 & f_8 = 0.55556 \\
x_9 = 1.9 & f_9 = 0.52632 \\
\hline
 & \text{Suma } 6.18773 \\
x_0 = 1.0 & \frac{1}{2}f_0 = 0.5 \\
x_{10} = 2.0 & \frac{1}{2}f_{10} = 0.25 \\
\hline
 & 6.93773 \cdot 1/10
\end{array}$$

$$\log_e 2 \approx 0.69377.$$

Puesto que la gráfica del integrando tiene su lado convexo hacia el eje x, este valor es demasiado grande.

Usando la regla de la tangente (7) se tiene

$$\begin{array}{ll}
x_0 + \frac{1}{2}h = 1.05 & f_{1/2} = 0.95238 \\
x_1 + \frac{1}{2}h = 1.15 & f_{3/2} = 0.86957 \\
x_2 + \frac{1}{2}h = 1.25 & f_{5/2} = 0.80000 \\
x_3 + \frac{1}{2}h = 1.35 & f_{7/2} = 0.74074 \\
x_4 + \frac{1}{2}h = 1.45 & f_{9/2} = 0.68966 \\
x_5 + \frac{1}{2}h = 1.55 & f_{11/2} = 0.64516 \\
x_6 + \frac{1}{2}h = 1.65 & f_{13/2} = 0.60606 \\
x_7 + \frac{1}{2}h = 1.75 & f_{15/2} = 0.57143 \\
x_8 + \frac{1}{2}h = 1.85 & f_{17/2} = 0.54054 \\
x_9 + \frac{1}{2}h = 1.95 & f_{19/2} = 0.51282 \\
\hline
 & 6.92836 \cdot 1/10
\end{array}$$

$$\log_e 2 \approx 0.69284,$$

valor que, debido a la convexidad de la curva, es demasiado pequeño.

Con la misma subdivisión se obtiene un resultado más preciso usando la regla de Simpson (8). Resulta

$x_1 = 1.1$	$f_1 = 0.90909$
$x_3 = 1.3$	$f_3 = 0.76923$
$x_5 = 1.5$	$f_5 = 0.66667$
$x_7 = 1.7$	$f_7 = 0.58824$
$x_9 = 1.9$	$f_9 = 0.52632$
	Suma $3.45955 \cdot 4$
	13.83820
$x_2 = 1.2$	$f_2 = 0.83333$
$x_4 = 1.4$	$f_4 = 0.71429$
$x_6 = 1.6$	$f_6 = 0.62500$
$x_8 = 1.8$	$f_8 = 0.55556$
	Suma $2.72818 \cdot 2$
	5.45636
	13.83820
$x_0 = 1.0$	$f_0 = 1.0$
$x_{10} = 2.0$	$f_{10} = 0.5$
	$20.79456 \cdot 1/30$

$$\log_e \approx 0.69315.$$

En realidad,

$$\log_e 2 = 0.693147\ldots.$$

6.2 Otros ejemplos de métodos numéricos

a. El "cálculo de errores"

El "cálculo de errores" es simplemente una aplicación numérica de un teorema básico del cálculo diferencial: una función $f(x)$, derivable un número suficiente de veces, puede ser representada en la vecindad de un punto por una función lineal con un error de orden superior al primero, por una función cuadrática con un error de orden superior al segundo, y así sucesivamente.

Considérese la aproximación lineal para la función $y = f(x)$. Si $y + \Delta y = f(x + \Delta x) = f(x + h)$, por el teorema de Taylor se tiene

$$\Delta y = hf'(x) + \frac{h^2}{2} f''(\xi),$$

donde $\xi = x + \theta h$ $(0 < \theta < 1)$ es un valor intermedio no necesariamente conocido en forma precisa. Si $h = \Delta x$ es pequeño, se obtiene la siguiente aproximación de interés práctico

$$\Delta y \approx hf'(x).$$

Así, el cociente incremental se reemplaza por la derivada, que lo aproxima, y el incremento de y se reemplaza por la aproximación dada por la expresión lineal en h.

Este sencillo resultado se usa para propósitos numéricos de la siguiente manera. Sean x e y dos cantidades físicas relacionadas por $y = f(x)$. Se pregunta entonces qué efecto tiene sobre la determinación de y una imprecisión en la medición de x. Si en vez del valor "verdadero" x se usa el valor aproximado $x + h$, el correspondiente valor de y difiere del valor verdadero $y = f(x)$ en la cantidad $\Delta y = f(x + h) - f(x)$. Por tanto, el error está dado aproximadamente por la relación anterior.

Mediante ejemplos se ilustra la utilidad de tales aproximaciones lineales.

Ejemplos. (*a*) Supóngase que en un triángulo ABC (ver Fig. 6.3) los lados b y c se determinan con precisión en tanto que el ángulo $\alpha = x$ se determina con un error $|\Delta x| < \delta$. ¿Cuál es el error correspondiente en el valor del tercer lado $y = a = \sqrt{b^2 + c^2 - 2bc \cos \alpha}$?

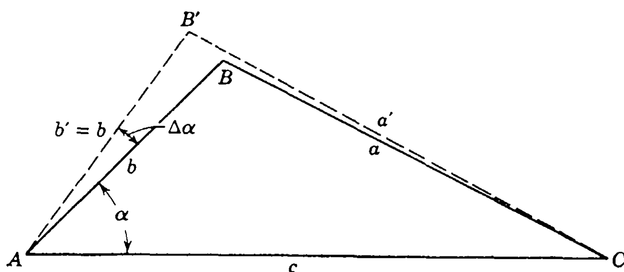


Figura 6.3

Se tiene $\Delta a \approx (bc \operatorname{sen} \alpha \, \Delta\alpha)/a$; y el porcentaje de error es por consiguiente

$$\frac{100\,\Delta a}{a} \approx \frac{100\,bc}{a^2} \operatorname{sen} \alpha \, \Delta\alpha.$$

En el caso especial en que $b = 400$ metros, $c = 500$ metros y $\alpha = 60°$, resulta $y = a = 458.2576$ metros, de manera que

$$\Delta a \approx \frac{200000}{458.2576} \times \tfrac{1}{2}\sqrt{3}\,\Delta\alpha.$$

Si $\Delta\alpha$ pudiera medirse con un error máximo de 10 segundos de arco, es decir, si

$$\Delta\alpha = 10'' = 4846 \times 10^{-8} \text{ radianes,}$$

se encontraría que en el caso más desfavorable

$$\Delta a \approx 1.83 \text{ cm.}$$

Así, el error sería a lo más del 0.004% aproximadamente.

(b) El ejemplo que sigue ilustra la utilidad de la "linealización" en problemas físicos.

Experimentalmente se sabe que si una varilla de metal tiene una longitud l_0 a la temperatura t_0, entonces a la temperatura t su longitud será $l = l_0(1 + \alpha(t - t_0))$, donde α depende sólo de t_0 y del material que compone la varilla. Ahora, si un reloj de péndulo conserva la hora exacta a la temperatura t_0, ¿cuántos segundos se atrasará diariamente si la temperatura se eleva a t_1?

Para el período de oscilación $T(l)$ se tiene (ver pp. 429-430)

$$T(l) = 2\pi\sqrt{\frac{l}{g}};$$

por tanto,

$$\frac{dT}{dl} = \frac{\pi}{\sqrt{lg}}.$$

Si el cambio en longitud es Δl, el cambio correspondiente en el período de oscilación es

$$\Delta T \approx \frac{\pi\,\Delta l}{\sqrt{l_0 g}},$$

donde $l_1 = l_0(1 + \alpha(t_1 - t_0))$ y $\Delta l = \alpha l_0(t_1 - t_0)$. Este es el tiempo de retraso por oscilación. El tiempo de retraso en un segundo es $\Delta T/T \approx \Delta l/2l_0$; por tanto, en un día el reloj se atrasa $43{,}200\,\Delta l/l_0 = 43{,}200\,\alpha(t_1 - t_0)$ segundos.

En éste y muchos otros casos en que la función considerada es un producto de varios factores, pueden simplificarse los cálculos tomando logaritmos en ambos miembros antes de derivar. Para el presente ejemplo se tiene:

$$\log T = \log 2\pi - \tfrac{1}{2}\log g + \tfrac{1}{2}\log l;$$

y al derivar,

$$\frac{1}{T}\frac{dT}{dl} = \frac{1}{2l}.$$

Reemplazando dT/dl por $\Delta T/\Delta l$, resulta

$$\frac{\Delta T}{T} = \frac{\Delta l}{2l},$$

lo cual coincide con el resultado anterior.

*b. Cálculo de π

Un ejemplo diferente, en el que se emplean recursos artificiales, es el siguiente. Es un ejemplo clásico, aunque quizá las máquinas calculadoras modernas lo han hecho obsoleto.

La serie de Leibnitz, $\pi/4 = 1 - 1/3 + 1/5 - 1/7 + \cdots$ [Ec. (7), sección 5.2, p. 464], obtenida a partir de la serie para la tangente inversa, no es adecuada para el cálculo de π debido a la extrema lentitud de su convergencia. Sin embargo, se puede calcular π con relativa facilidad mediante el artificio siguiente. Si en el teorema de adición para la tangente,

$$\tan(\alpha + \beta) = \frac{\tan \alpha \tan \beta}{1 - \tan \alpha \tan \beta},$$

se introducen las funciones inversas $\alpha = \arctan u$, $\beta = \arctan v$, se obtiene la fórmula

$$\arctan u + \arctan v = \arctan\left(\frac{u + v}{1 - uv}\right).$$

Ahora bien, escogiendo u y v de manera que $(u + v)/(1 - uv) = 1$, se obtiene el valor $\pi/4$ en el miembro derecho y, si u y v son números pequeños, puede fácilmente calcularse el miembro izquierdo por medio de series conocidas. Por ejemplo, si se hacen $u = 1/2$, $v = 1/3$, como hizo Euler, se obtiene

$$\frac{\pi}{4} = \arctan 1/2 + \arctan 1/3. \tag{9}$$

Si, además, se observa que $\dfrac{1/3 + 1/7}{1 - 1/21} = \dfrac{1}{2}$,

se tiene $\arctan 1/2 = \arctan 1/3 + \arctan 1/7$, de manera que, por (9),

$$\frac{\pi}{4} = \arctan 1/3 + \arctan 1/7.$$

Usando esta fórmula Vega calculó el número π con 140 cifras decimales.

Por medio de la ecuación $(1/5 + 1/8)/(1 - 1/40) = 1/3$, se obtiene además

$$\arctan 1/3 = \arctan 1/5 + \arctan 1/8$$

en donde:

$$\frac{\pi}{4} = 2 \arctan 1/5 + \arctan 1/7 + 2 \arctan 1/8.$$

Este desarrollo es sumamente útil para el cálculo de π por medio de la serie $\arctan x = x - x^3/3 + x^5/5 - \cdots$; pues si se substituye x por uno de los valores $1/5$, $1/7$ ó $1/8$, se obtiene con pocos términos un alto grado de precisión ya que éstos disminuyen rápidamente.

El lector que no tenga especial interés en estas manipulaciones, que requieren habilidad aunque sean artificiosas, puede conformarse con entender el principio.

*c. Cálculo de logaritmos

Para el cálculo numérico de logaritmos la serie logarítmica [Ec. (5), p. 463]

$$\tfrac{1}{2} \log \frac{1 + x}{1 - x} = x + \frac{x^3}{3} + \frac{x^5}{5} + \cdots,$$

donde $0 < x < 1$, mediante la substitución

$$\frac{1+x}{1-x} = \frac{p^2}{p^2-1}, \qquad x = \frac{1}{2p^2-1},$$

se transforma en la serie

$$\log p = \tfrac{1}{2}\log(p-1) + \tfrac{1}{2}\log(p+1) + \frac{1}{2p^2-1} + \frac{1}{3(2p^2-1)^3} + \cdots$$

donde $2p^2 - 1 > 1$ o $p^2 > 1$. Si p es un entero y $p + 1$ puede descomponerse en factores enteros más pequeños (por ejemplo, si $p + 1$ es par), esta última serie expresa el logaritmo de p mediante los logaritmos de enteros menores, más una serie cuyos términos disminuyen muy rápidamente y cuya suma puede por tanto calcularse con suficiente precisión usando sólo unos cuantos términos. A partir de esta serie pueden entonces calcularse sucesivamente los logaritmos de cualesquier números primos y, en consecuencia, de cualquier número, suponiendo que ya se haya calculado el valor log 2 (por ejemplo, mediante su representación integral, como en las pp. 508-509).

La exactitud de esta determinación de log p puede estimarse más fácilmente por medio de la serie geométrica que usando la fórmula general del residuo. Para el residuo R_n de la serie, esto es, la suma de todos los términos que siguen al término $1/n(2p^2-1)^n$, se tiene

$$R_n < \frac{1}{(n+2)(2p^2-1)^{n+2}}\left(1 + \frac{1}{(2p^2-1)^2} + \frac{1}{(2p^2-1)^4} + \cdots\right)$$
$$= \frac{1}{(n+2)(2p^2-1)^n} \cdot \frac{1}{(2p^2-1)^2 - 1},$$

y esta fórmula da inmediatamente la estimación requerida para el error.

Calculemos, por ejemplo, $\log_e 7$ (con la hipótesis de que log 2 y log 3 ya han sido determinados numéricamente), usando los cuatro primeros términos de la serie Tenemos

$$p = 7, \qquad 2p^2 - 1 = 97,$$

$$\log 7 = 2\log 2 + \tfrac{1}{2}\log 3 + \frac{1}{97} + \frac{1}{3\cdot 97^3} + \cdots;$$

$$1/97 \approx 0.01030928, \qquad \frac{1}{3\cdot 97^3} \approx 0.00000037,$$

$$2\log 2 \approx 1.38629436, \qquad \tfrac{1}{2}\log 3 \approx 0.54930614;$$

y, por tanto,

$$\log_e 7 \approx 1.94591015.$$

La estimación del error es

$$R_n < \frac{1}{5 \cdot 97^3} \times \frac{1}{97^2 - 1} < \frac{1}{36 \times 10^9}.$$

Sin embargo, obsérvese que cada uno de los cuatro números que se han sumado fue dado con un error de 5×10^{-9}, de manera que la última cifra decimal en el valor calculado de log 7 podría estar afectada por un error de ± 2. De hecho, sin embargo, esta última cifra es también correcta.

6.3 Solución numérica de ecuaciones

Se agregan aquí algunas observaciones sobre la solución numérica de la ecuación $f(x) = 0$, donde $f(x)$ no es necesariamente un polinomio.[1] Se comienza con un primer valor de prueba x_0 para una de las raíces y luego se afina esta aproximación. Queda abierta la cuestión de cómo se escoge el primer valor de aproximación para la raíz y cuán buena es esta aproximación. Puede, por ejemplo, hacerse una conjetura o, mejor aún, obtenerse una primera aproximación a partir de la gráfica de la función $y = f(x)$, cuya intersección con el eje x indica la raíz deseada.

Luego se intenta afinar la aproximación mediante un proceso o transformación que lleve el valor x_0 a una "segunda aproximación", y se repite este proceso. Resolver numéricamente la ecuación $f(x) = 0$ consiste en llevar a cabo esas aproximaciones sucesivas en forma repetida (o, como se dice, "iterar" el proceso) con la esperanza de que los valores iterados $x_1, x_2, \ldots, x_n$ converjan satisfactoriamente a la raíz ξ. A continuación se consideran dichos procedimientos y se discute su precisión.

a. Método de Newton

Descripción del método. El procedimiento iterativo de Newton se basa en un principio fundamental del cálculo diferencial: el reemplazo de una curva por una tangente en la inmediata vecindad del punto de contacto. Partiendo con un primer valor aproximado, x_0, para una raíz ξ de la ecuación $f(x) = 0$, se considera el punto de la gráfica de la función $y = f(x)$ cuyas coordenadas son $x = x_0$, $y = f(x_0)$. Con el fin de encontrar una mejor aproximación para la intersección ξ de la curva con el eje x, se determina el punto x_1 en el cual la tangente en $x = x_0$, $y = f(x_0)$ intersecta al eje x. La abscisa x_1 de esta intersección representa una nueva y, en ciertas circunstancias, mejor aproximación que x_0 para la raíz deseada, ξ, de la ecuación.

[1] Estamos, por supuesto, interesados solamente en la determinación de raíces reales de $f(x) = 0$.

De la figura 6.4 se obtiene de inmediato

$$\frac{f(x_0)}{x_0 - x_1} = f'(x_0);$$

y, por tanto, la nueva aproximación

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}. \tag{10}$$

Partiendo con x_1 como aproximación, se repite el proceso para encontrar $x_2 = x_1 - f(x_1)/f'(x_1)$, y así sucesivamente.

La utilidad de este procedimiento depende esencialmente de la naturaleza de la curva. En la situación que se indica en la Fig. 6.4, las aproximaciones sucesivas x_n convergen con precisión creciente a la raíz deseada ξ.

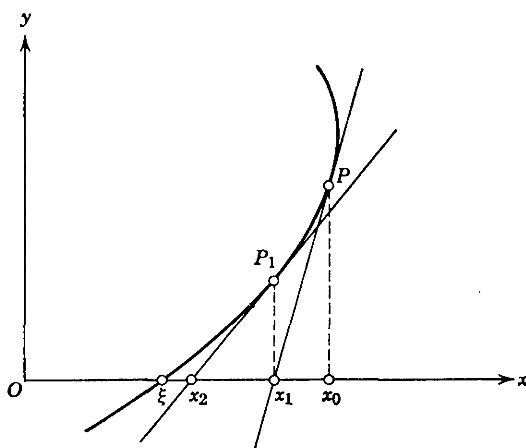


Figura 6.4 Método de aproximación de Newton.

Sin embargo, la Fig. 6.5 muestra que, con una elección plausible del valor original x_0, nuestra construcción no necesariamente converge a la raíz requerida. Es necesario, por consiguiente, examinar en general las circunstancias bajo las cuales el método de Newton proporciona aproximaciones útiles a la solución de la ecuación.

Convergencia cuadrática del método de Newton

Suponiendo que en un intervalo suficientemente amplio en torno a la raíz ξ la segunda derivada $f''(x)$ no es "demasiado grande" y la primera derivada $f'(x)$ no "demasiado pequeña, el hecho principal relativo a la aproximación de Newton es que los "errores" sucesivos

$$h_1 = \xi - x_1, \quad h_2 = \xi - x_2, \ldots, h_n = \xi - x_n, \ldots$$

convergen cuadráticamente a cero en el sentido de que $|h_{n+1}| \leq \mu h_n^2$ para una constante fija μ. Lo anterior indica una razón de convergencia extremadamente rápida. La desigualdad, escrita en la forma $|h_{n+1}\mu| \leq |h_n\mu|^2$, implica, por ejemplo, que cuando $|h_n\mu| < 10^{-m}$ se tiene $|h_{n+1}\mu| < 10^{-2m}$, o sea, el número de "dígitos significativos" en μx_n se duplica en cada etapa.

La demostración de la convergencia cuadrática es inmediata. De las relaciones $x_{n+1} = x_n - f(x_n)/f'(x_n)$ y $f(\xi) = 0$, se encuentra que

$$h_{n+1} = \xi - x_{+1} = \xi - x_n - \frac{f(\xi) - f(x_n)}{f'(x_n)}.$$

Por la fórmula de Taylor,

$$f(\xi) - f(x_n) = (\xi - x_n)f'(x_n) + \tfrac{1}{2}(\xi - x_n)^2 f''(\eta),$$

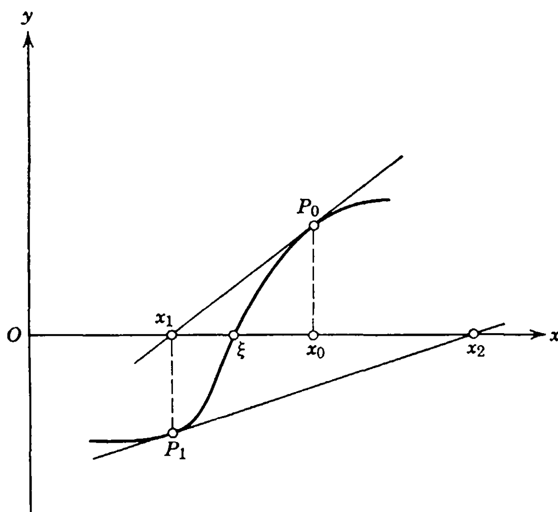


Figura 6.5

con η comprendido entre ξ y x_n. Por tanto,

$$(11) \qquad h_{n+1} = -\frac{f''(\eta)}{2f'(x_n)}h_n^2 .$$

Para establecer la convergencia, supóngase que x_n pertenece ya a un intervalo fijo $\xi - \delta < x < \xi + \delta$ en el cual $|f''|$ tiene el valor máximo M_2, $|f'|$ el valor mínimo m_1 (positivo), y δ es tan pequeño que $\frac{1}{2}\delta M_2/m_1 < 1$. Haciendo $\mu = \frac{1}{2}M_2/m_1$, se tiene $\mu\delta < 1$ y

$$|h_{n+1}| \leq \mu\,|h_n|^2 \leq \mu\delta\,|h_n| < |h_n|.$$

Esta desigualdad muestra, en primer lugar, que x_{n+1} pertenece nuevamente a la misma vecindad δ de ξ, de manera que el argumento puede repetirse.

De esta manera, si x_0 está en la vecindad δ de ξ, lo mismo ocurrirá para todos los x_n subsiguientes. De $|h_{n+1}| \leq \mu\delta\,|h_n|$ se deduce entonces que $|h_{n+1}| \leq (\mu\delta)^{n+1}\,|h_0|$, lo cual implica que $h_n \to 0$, o sea, que $x_n \to \xi$. Además, la ley del decrecimiento cuadrático, $|h_{n+1}| \leq \mu\,|h_n|^2$, se cumplirá para los errores. Resulta entonces evidente que el método de Newton proporciona una sucesión x_n que efectivamente converge a la solución ξ siempre que f' y f'' existan y sean continuas cerca de ξ, que $f'(\xi) \neq 0$ y que x_0 esté suficientemente cerca de la solución. El carácter cuadrático de la aproximación es a menudo una ventaja decisiva del método de Newton respecto a otros (ver p. 522).

*b. La Regla de la posición falsa

El método de Newton es el caso límite de un método más antiguo, el de la "regla de la posición falsa", en el cual aparece la secante en lugar de la tangente. Supongamos que se conocen dos puntos (x_0, y_0) y (x_1, y_1) en la vecindad de la intersección con el eje x requerida. Si reemplazamos la curva por la secante que une estos dos puntos, la intersección de ésta con el eje x puede ser una mejor aproximación a la raíz[1] requerida de la ecuación. Para la abscisa ξ del punto de intersección se tiene (Fig. 6.6)

$$(12) \qquad \frac{\xi - x_0}{f(x_0)} = \frac{\xi - x_1}{f(x_1)},$$

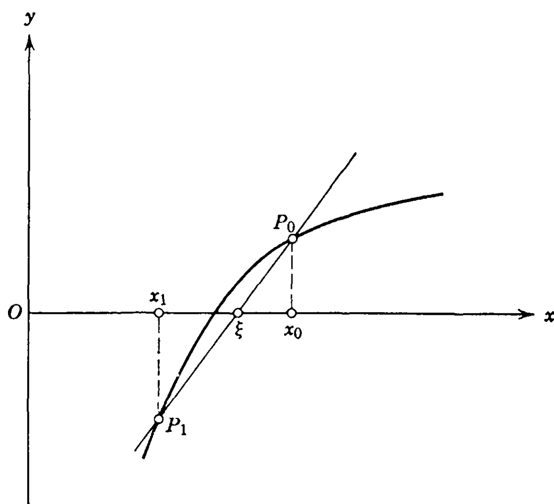


Figura 6.6 La regla de la posición falsa.

[1] Esto equivale esencialmente a aplicar la interpolación lineal a la función inversa.

lo cual conduce a

$$\xi = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)}$$
$$= \frac{x_0 f(x_1) - x_0 f(x_0) + x_0 f(x_0) - x_1 f(x_0)}{f(x_1) - f(x_0)};$$

o sea,

$$(13) \qquad \xi = x_0 - \frac{f(x_0)}{\dfrac{f(x_1) - f(x_0)}{x_1 - x_0}}.$$

Esta fórmula, que determina la aproximación ξ con respecto a x_0 y x_1, constituye la regla de la 'posición falsa". Es útil cuando uno de los valores de la función es positivo y el otro es negativo, como en la Fig. 6.6, por ejemplo, en donde $y_0 > 0$ y $y_1 < 0$.

La fórmula de aproximación de Newton resulta como un caso límite cuando $x_1 \to x_0$, pues el denominador del segundo término del miembro derecho de la fórmula (13) tiende a $f'(x_0)$ conforme x_1 tiende a x_0.

Aunque la regla de la posición falsa puede considerarse como más elemental que el método de Newton, este último tiene la gran ventaja de requerir como aproximación inicial un solo valor de x en vez de dos.

c. El método iterativo

El esquema de iteración. Consideremos ahora un esquema de grandes posibilidades en la solución de ecuaciones escritas en la forma

$$x = \phi(x),$$

donde ϕ es una función continua con una derivada también continua. La solución de ecuaciones de la forma $f(x) = 0$ puede reducirse a la de $x = \phi(x)$ haciendo $\phi(x) = x - c(x)f(x)$, donde $c(x)$ es cualquier función diferente de cero.

En el *método iterativo*,[1] particularmente sugerente, se empieza también con un valor de aproximación inicial x_0 escogido adecuadamente y luego se determina una sucesión de valores $x_1, x_2, x_3 \ldots$, por medio de las condiciones

$$x_{n+1} = \phi(x_n), \qquad n = 0, 1, 2, \ldots.$$

Si esta sucesión "de iteración", x_n, converge a un límite ξ, entonces $\xi = \phi(\xi)$ es una solución de nuestra ecuación, pues en ese caso $\lim_{n\to\infty} x_{n+1} = \xi$ y $\lim_{n\to\infty} \phi(x_n) = \phi(\xi)$ debido a la continuidad de la función ϕ.

---

[1] Llamado algunas veces "método de aproximaciones sucesivas". El método es usado en muchos contextos matemáticos diferentes para la solución de ecuaciones de una clase u otra.

Convergencia. La sucesión de valores x_n en el proceso iterativo converge, con una hipótesis muy general, a una solución: Si la primera aproximación x_n está en un intervalo[2] J en torno a ξ, en el cual

$$|\phi'(x)| < q$$

con una constante $q < 1$, entonces x_n converge a ξ.

Puesto que, por hipótesis, x_0 está en J, se tiene

$$x_1 - \xi = \phi(x_0) - \phi(\xi).$$

El miembro derecho de esta ecuación, es por el teorema del valor medio, igual a $(x_0 - \xi)\phi'(\bar{x})$, siendo $\bar{x}$ un punto de J. Así, por nuestras hipótesis es

$$|x_1 - \xi| \leq q\,|x_0 - \xi|,$$

de manera que x_1 pertenece a J y, por tanto, también es

$$|x_2 - \xi| \leq q\,|x_1 - \xi| \leq q^2\,|x_0 - \xi|.$$

En general se obtiene

$$|x_n - \xi| \leq q^n\,|x_0 - \xi|;$$

y, como $q^n \to 0$ conforme $n \to \infty$, nuestro aserto queda demostrado.

De lo anterior se ve, además, que la sucesión de iteración x_n no converge cuando $\phi'(x) > 1$ en un intervalo que contiene a ξ; y si $|\phi'(\xi)| = 1$ no es posible hacer una afirmación de validez general.

Puntos fijos atractivos y repulsivos

Es útil considerar el proceso de iteración en términos de transformaciones. La función $y = \phi(x)$ representa una transformación que proyecta un punto x del eje numérico sobre un punto imagen y de dicho eje (ver p. 44). La solución ξ es entonces un punto que no cambia bajo la transforción ϕ, denominado *punto fijo,* y el problema en este caso consiste en encontrar un punto fijo de la transformación. Este problema puede resolverse mediante la iteración siempre que $|\phi'(\xi)| \leq q < 1$, como ya se ha visto.

La transformación $y = \phi(x)$ en la vecindad de la raíz o punto fijo ξ tiene, para $|\phi'(x)| < q < 1$, la propiedad de ser una *contracción,* esto es, de disminuir la distancia del punto original al punto fijo. Los puntos fijos de transformaciones de contracción se llaman *puntos fijos atractivos.* Su construcción mediante iteraciones converge como los términos de una serie geométrica de cociente q.

Si la raíz ξ o el correspondiente punto fijo de nuestra transformación está en un intervalo en el cual $|\phi'(x)| > r$, donde r es una constante ma-

---

[2] Aunque se desconoce ξ, a menudo es posible determinar *a priori* tal intervalo.

yor que 1, la *transformación es de expansión,* el proceso de iteración diverge y el *punto fijo* se denomina *repulsivo.*

Si en el punto fijo se tiene $|\phi'(\xi)| = 1$, no es posible hacer afirmaciones generales relativas a la convergencia de las iteraciones. Tales puntos fijos se denominan a veces *indiferentes.*

Debe recalcarse lo siguiente: un punto fijo ξ de una transformación ϕ es automáticamente también un punto fijo para ψ, la transformación inversa: $\xi = \psi(\xi)$. Si $|\phi'(\xi)| > 1$ en la vecindad de una raíz ξ y $x = \psi(y)$ es la función inversa de ϕ, entonces $|\psi'(\xi)| < 1$. Así, ξ es un punto fijo atractivo para esta transformación inversa, y es posible reemplazar el esquema de iteración originalmente divergente por uno convergente para la transformación inversa. Como un ejemplo, considérese la ecuación

$$x = \tan x.$$

De las gráficas de las funciones $y = x$ e $y = \tan x$ resulta claro que éstas se intersectan en algún punto del intervalo $\pi < x < \frac{3}{2}\pi$ y que la ecuación tendrá un raíz ξ en ese intervalo (Fig. 6.7). Puesto que

$$\frac{d \tan x}{dx} = \frac{1}{\cos^2 x} > 1,$$

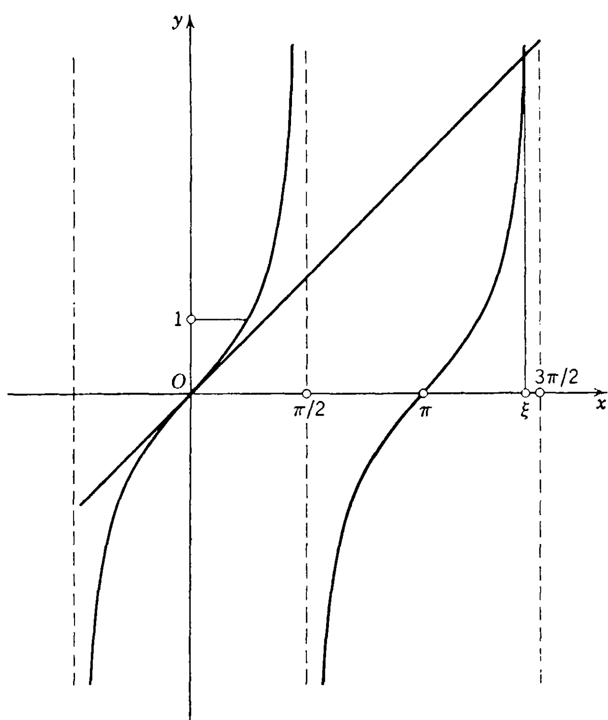


Figura 6.7 Intersección (ξ, ξ) de las curvas $y = \tan x$ y $y = x$.

el procedimiento de iteración con un punto cualquiera x_0 del intervalo no converge. Sin embargo, se obtiene una sucesión de iteración convergente si se escribe la ecuación en la forma inversa (usando la notación arc tan x para la rama principal),

$$x = \arctan x + \pi.$$

Puesto que aquí

$$\frac{d}{dx} \arctan x = \frac{1}{1 + x^2} < 1,$$

la sucesión definida por $x_{n+1} = \arctan x_n + \pi$ y, digamos, $x_0 = \pi$, converge a ξ.

d. Iteraciones y procedimiento de Newton

Como ya se ha mencionado, el problema de resolver una ecuación de la forma $f(x) = 0$ puede reducirse al de resolver la ecuación $x = \phi(x)$ si se escoge para ϕ cualquier expresión de la forma

$$\phi(x) = x - c(x)f(x),$$

donde $c(x)$ es una función que no se anula. Si se desea resolver la ecuación resultante $x = \phi(x)$ por iteración es necesario asegurar, escogiendo $c(x)$ adecuadamente, que el punto fijo ξ de la transformación ϕ es "atractivo", esto es, que $|\phi'(\xi)| < 1$. En estas condiciones, se tiene para la solución ξ de $f(\xi) = 0$:

$$\phi'(\xi) = 1 - c'(\xi)f(\xi) - c(\xi)f'(\xi) = 1 - c(\xi)f'(\xi).$$

Lo más sencillo es tomar para $c(x)$ la expresión $1/f'(x)$. Entonces, en efecto, $|\phi'(\xi)| = 0 < 1$. Esta elección de $c(x)$ conduce a la sucesión de iteración

$$x_{n+1} = \phi(x_n) = x_n - \frac{f(x_n)}{f'(x_n)},$$

que es precisamente la sucesión de aproximaciones (10) (pp. 514-515) del método de Newton. Para el error $x_n - \xi = h_n$ se tiene la estimación

$$|h_{n+1}| = |\phi(x_n) - \phi(\xi)| \leq q h_n,$$

donde q es el máximo de $|\phi'(x)|$ en el intervalo cuyos puntos extremos son ξ y x_n. Puesto que en este caso

$$\phi'(x) = \frac{f(x)f''(x)}{f'^2(x)}$$

y $f(x) = f(x) - f(\xi) = f'(\eta)(x - \xi)$, se ve que q mismo es del orden de h_n, confirmándose así nuevamente el carácter cuadrático de la aproximación en el método de Newton.

Otra elección sencilla para $c(x)$ la constituye el valor constante $1/f'(x_0)$, lo cual conduce a la fórmula de recurrencia

$$x_{n+1} = \phi(x_n) = x_n - \frac{f(x_0)}{f'(x_0)}.$$

Aquí es $\phi'(\xi) = 1 - f'(\xi)/f'(x_0)$. Si f' es continua y diferente de cero se tendrá un punto fijo atractivo ξ cuando la aproximación inicial x_0 esté tan cerca de la solución ξ que sea

$$|\phi'(\xi)| = \frac{|f'(x_0) - f'(\xi)|}{|f'(x_0)|} < 1.$$

Esta sucesión de iteración es algo más simple que la usada en el método de Newton; sin embargo, su convergencia será más lenta (parecida a la de una serie geométrica), como ocurre con la mayor parte de los esquemas de iteración.

Ejemplos. Como un ejemplo, considérese la ecuación cúbica

$$f(x) = x^3 - 2x - 5 = 0.$$

Puesto que $f(2) = -1 < 0$, $f(3) = 16 > 0$, es claro que existe una raíz en el intervalo $2 < x < 3$. Además, como $f'(x) = 3x^2 - 2 > 3(2)^2 - 2 > 0$, ese intervalo contiene sólo una raíz. Usando el método de Newton con la aproximación inicial $x_0 = 2$ se encuentra sucesivamente

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 2 - \frac{-1}{3(2)^2 - 2} = 2.1, \qquad f(x_1) = 0.061,$$

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = 2.1 - \frac{0.061}{3(2.1)^2 - 2} = 2.094568.$$

Análogamente, puesto que $f(2.1) > 0$, $f(2) < 0$, la raíz ξ está entre 2 y 2.1. En el intervalo $1.9 < x < 2.2$ y, por consiguiente, también en el intervalo $\xi - 0.1 < x < \xi + 0.1$, se tienen las estimaciones

$$|f''(x)| = |6x| < 6(2.2) = 13.2,$$
$$f'(x) = 3x^2 - 2 > 3(1.9)^2 - 2 = 8.83.$$

Se deduce entonces [ver (11), p. 516] que

$$|\xi - x_{n+1}| \leq \frac{13.2}{2(10.83)} |x_n - \xi|^2 < 0.75\, |x_n - \xi|^2;$$

en el supuesto de que $|x_n - \xi| < 0.1$. Como $|x_0 - \xi| = |\xi - 2| < 0.1$, se obtiene sucesivamente

$$|x_1 - \xi| < (0.75)(0.1)^2 = 0.0061,$$
$$|x_2 - \xi| < (0.61)(0.0061)^2 < 0.000042.$$

Si este grado de aproximación no es suficiente, puede obtenerse una aproximación posterior, x_3, con un error $< (0.75)(0.000042)^2 < 0.000\,000\,001\,3$.

Todo x_n posterior a x_0 debe ser mayor que ξ, como se desprende del hecho de que f' y f'' son positivas, lo cual implica que

$$h_{n+1} = -f''(\eta)h_n^2/2f'(x_n) < 0.$$

Por otra parte, la aplicación de la regla de la posición falsa [(13, p. 518] a los valores x_0, x_1 da para la intersección del eje x con la secante que une los puntos $(x_0, f(x_0))$ y $(x_1, f(x_1))$ el resultado

$$\xi = x_0 - \frac{f(x_0)(x_1 - x_0)}{f(x_1) - f(x_0)} = 2.09425\cdots.$$

Puesto que la curva es convexa en el intervalo considerado, la secante queda por encima de la curva y la aproximación de ξ debe ser menor que la raíz ξ.

Como un segundo ejemplo, resolvamos la ecuación

$$f(x) = x \log_{10} x - 2 = 0.$$

Tenemos $f(3) = -0.6$ y $f(4) = +0.4$, y por consiguiente usamos $x_0 = 3.5$ como primera aproximación. Usando tablas logarítmicas de diez cifras decimales obtenemos las siguientes aproximaciones sucesivas

$$x_0 = 3.5, \qquad x_1 = 3.598,$$
$$x_2 = 3.5972849, \qquad x_3 = 3.5972850235.$$

Apéndice

*A.1 Fórmula de Stirling

En muchas aplicaciones, particularmente en Estadística y en la Teoría de Probabilidades, se presenta la necesidad de disponer de una aproximación sencilla para $n!$ en la forma de una función elemental de n. Tal expresión está dada por el siguiente teorema, que lleva el nombre de su descubridor, Stirling (ver también capítulo 8, p. 630).

Conforme $n \to \infty$,

(14) $$\frac{n!}{\sqrt{2\pi}\, n^{n+1/2}e^{-n}} \to 1;$$

o, en forma más precisa,

(14a) $$\sqrt{2\pi}\, n^{n+1/2}e^{-n} < n! < \sqrt{2\pi} n^{n+1/2}e^{-n}\left(1 + \frac{1}{4n}\right)$$

En otras palabras, las expresiones $n!$ y $\sqrt{\pi}\, n^{n+1/2}e^{-n}$ difieren solamente por un porcentaje pequeño cuando el valor de n es grande —o, como usualmente se dice, las dos expresiones son *asintóticamente iguales*— y, al mismo tiempo, el factor $1 + 1/4n$ proporciona una estimación del grado de exactitud de la aproximación.

Esta notable fórmula aparece cuando se trata de evaluar el área bajo la curva $y = \log x$.[1] Por integración (p. 297) se encuentra que A_n, el área exacta bajo esta curva entre las abscisas $x = 1$ y $x = n$, está dada por

$$(15) \qquad A_n = \int_1^n \log x\, dx = x \log x - x \Big|_1^n = n \log n - n + 1.$$

Sin embargo, si se estima el área usando la regla del trapezoide, trazando verticales por los puntos $x = 1,\ x = 2, \ldots, x = n$ como en la Fig. 6.8, se obtiene un valor aproximado T_n para el área [ver (6), p. 503]:

$$(16) \qquad \begin{aligned} T_n &= \log 2 + \log 3 + \cdots + \log (n-1) + \tfrac{1}{2}\log n \\ &= \log n! - \tfrac{1}{2}\log n. \end{aligned}$$

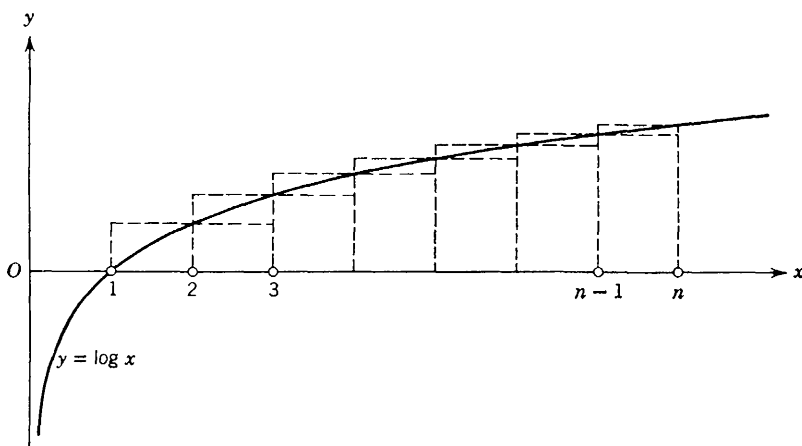


Figura 6.8

Haciendo la hipótesis razonable de que A_n y T_n son del mismo orden de magnitud, se encuentra que $n!$ es del mismo orden de magnitud que $n^{n+1/2}e^{-n}$, lo cual constituye la esencia de lo expresado por la fórmula de Stirling.

Para precisar este argumento demostraremos primero que la diferencia $a_n = A_n - T_n$ es acotada, de lo cual se desprenderá inmediatamente que $T_n = A_n(1 - a_n/A_n)$ es del mismo orden de magnitud que A_n. La can-

[1] El método empleado aquí es un caso especial de la fórmula de Euler MacLaurin que será discutida en el capítulo 8, p. 642.

tidad $a_{k+1} - a_k$ es la diferencia entre el área bajo la curva y el área bajo la secante, en la franja $k < x < k + 1$. Puesto que la curva es cóncava y queda por encima de la secante, $a_{k+1} - a_k$ es positivo y $a_n = (a_n - a_{n-1}) + (a_{n-1} - a_{n-2}) + \cdots + (a_2 - a_1) + a_1$, monótona creciente. Además, la diferencia $a_{k+1} - a_k$ es obviamente menor (ver Fig. 6.9) que la diferencia

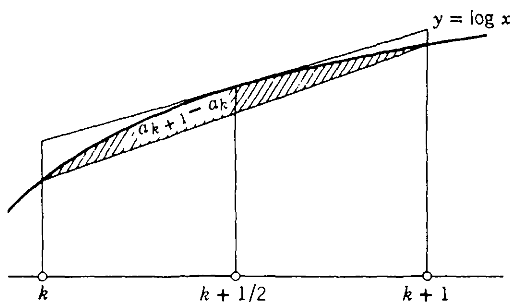


Figura 6.9

entre el área bajo la tangente en $x = k + \frac{1}{2}$ y el área bajo la secante. Por tanto, se tiene la desigualdad

$$\begin{aligned} a_{k+1} - a_k &< \log(k + \tfrac{1}{2}) - \tfrac{1}{2}\log k - \tfrac{1}{2}\log(k+1) \\ &= \tfrac{1}{2}\log\left(1 + \frac{1}{2k}\right) - \tfrac{1}{2}\log\left[1 + \frac{1}{2(k+\frac{1}{2})}\right] \\ &< \tfrac{1}{2}\log\left(1 + \frac{1}{2k}\right) - \tfrac{1}{2}\log\left[1 + \frac{1}{2(k+1)}\right]. \end{aligned}$$

Sumando estas desigualdades para $k = 1,2, \ldots, n - 1$, se encuentra que todos los términos del lado derecho, excepto dos, se cancelan y (puesto que $a_1 = 0$) resulta

$$a_n < \tfrac{1}{2}\log \tfrac{3}{2} - \tfrac{1}{2}\log\left(1 + \frac{1}{2n}\right) < \tfrac{1}{2}\log \tfrac{3}{2}.$$

Como a_n, además de ser acotada, es monótona creciente, tiende a un límite a conforme $n \to \infty$. La desigualdad para $a_{k+1} - a_k$ da ahora

$$a - a_n = \sum_{k=n}^{\infty} (a_{k+1} - a_k) < \tfrac{1}{2}\log\left(1 + \frac{1}{2n}\right).$$

Ya que, por definición, $A_n - T_n = a_n$, por (15) y (16) se tiene

$$\log n! = 1 - a_n + (n + \tfrac{1}{2})\log n - n,$$

o bien, escribiendo $\alpha_n = e^{1-a_n}$,

$$n! = \alpha_n n^{n+1/2} e^{-n}.$$

La sucesión α_n es monótona *decreciente* y tiende al límite $\alpha = e^{1-a}$; y, por tanto,

$$1 < \frac{\alpha_n}{\alpha} = e^{a-a_n} < e^{(1/2)\log(1+1/2n)}$$
$$= \sqrt{1 + \frac{1}{2n}} < 1 + \frac{1}{4n}.$$

Se tiene, por consiguiente,

$$\alpha n^{n+1/2}e^{-n} < n! < \alpha n^{n+1/2}e^{-n}\left(1 + \frac{1}{4n}\right).$$

Sólo nos resta encontrar el valor verdadero del límite a. Para ello usaremos la fórmula (80) del capítulo 3, p. 302:

$$\sqrt{\pi} = \lim_{n\to\infty} \frac{(n!)^2 2^{2n}}{(2n)!\sqrt{n}}.$$

Reemplazando $n!$ por $\alpha_n n^{n+1/2}e^{-n}$ y $(2n)!$ por $\alpha_{2n}2^{2n+1/2}n^{2n+1/2}e^{-2n}$ se obtiene inmediatamente

$$\sqrt{\pi} = \lim_{n\to\infty} \frac{\alpha_n{}^2}{\alpha_{2n}\sqrt{2}}$$
$$= \frac{\alpha^2}{\alpha\sqrt{2}},$$

por lo que $\alpha = \sqrt{2\pi}$. La demostración de la fórmula de Stirling queda así concluida.

Además de su interés teórico, la fórmula de Stirling es una herramienta de gran utilidad para el cálculo numérico de $n!$ cuando n es grande. En vez de multiplicar un gran número de enteros, sólo tiene que calcularse la expresión de Stirling por medio de logaritmos, lo que supone un número de operaciones considerablemente menor. Así, para $n = 10$ se obtiene el valor 3598696 mediante la expresión de Stirling (usando tablas de siete cifras), en tanto que el valor exacto de 10! es 3628800. El porcentaje de error es apenas de $\frac{5}{6}\%$.

PROBLEMAS

SECCION 6.1, página 502

1. Si $f''(x) \geq 0$, demostrar que la regla del trapezoide conduce a un valor mayor y la regla de la tangente a uno menor que el valor exacto de la integral de f.

2. Estimar el valor $h = (b - a)/n$ necesario para calcular hasta la p-ésima cifra decimal, usando la regla de Simpson, el valor de

(*a*) $\log 2 = \int_1^2 \frac{1}{x}\,dx,$ (*b*) $\pi = 4\int_0^1 \frac{1}{1+x^2}\,dx.$

3. Estimar en términos de k y s ($k < 1$ y $s < 1$) el número de puntos necesarios para calcular con un error menor de ε el valor de la integral elíptica

$$u(s) = \int_0^s \frac{dx}{\sqrt{(1 - x^2)(1 - k^2x^2)}}.$$

4. Sea $f(x)$ una función continua en el intervalo $\alpha \leq x \leq \alpha + h$, con una derivada acotada uniformemente: $|f'(x)| \leq M_1$ para una constante M_1. Demostrar que para cualquier punto fijo ξ ($\alpha \leq \xi \leq \alpha + h$) se tiene la estimación

$$\left| \int_\alpha^{\alpha+h} f(x)\, dx - hf(\xi) \right| \leq \frac{M_1h^2}{2}$$

5. Calcular numéricamente $\int_\alpha^\infty e^{-x^2}\, dx$ con un error menor que 1/100.

SECCION 6.2, página 509

1. El período de un péndulo está dado por

$$T = 2\pi \sqrt{\frac{l}{g}}$$

donde l es la longitud del péndulo. Si dicho péndulo acciona un reloj que se adelanta un minuto por día, determínese la corrección necesaria en l.

2. Para medir la altura de una colina, se observa desde la planicie una torre de 100 metros situada en la cima. El ángulo de elevación de la base de la torre es de 42° y la torre misma subtiende un ángulo de 6°. ¿Cuáles son los límites del error en la determinación de la altura si el ángulo de 42° está sujeto a un error de 1°?

SECCION 6.3, página 514

1. (a) Para resolver la ecuación $x = f(x)$ muéstrese la mejor forma de escoger la constante a de manera que el esquema de iteración

$$x_{k+1} = x_k + a[x_k - f(x_k)]$$

converja lo más rápidamente posible en la vecindad de la solución.

(b) Aplíquese este método con el fin de resolver la ecuación $\sqrt{A}$,

$$x = \frac{A}{x}.$$

(c) Demostrar que si $A \geq 1$, el número de cifras decimales exactas por lo menos se duplica en cada paso del esquema de iteración obtenido en (b).

2. (a) Muéstrese la mejor forma de elegir un polinomio

$$g(x) = a + bx^2$$

de manera tal que el esquema de iteración para $\sqrt{A}$,

$$x_{k+1} = x_k + g(x_k)\left(x_k - \frac{A}{x_k}\right),$$

converja lo más rápidamente posible en la vecindad de la solución.

(b) Estimar la rapidez de la convergencia.

(c) Mostrar cómo mejorar posteriormente la convergencia mediante la elección adecuada de polinomios $g(x)$ de grado superior.

3. Investigar la posibilidad de esquemas adecuados del tipo tratado en los problemas 1 y 2, para el cálculo de $\sqrt[3]{A}$.

SECCION A.1, página 523

1. Probar que $\lim\limits_{n\to\infty} \dfrac{\sqrt[n]{n!}}{n} = \dfrac{1}{e}$.

***2.** Considerando $\displaystyle\int_{1/2}^{n+1/2} \log\,(\alpha + x)\,dx$, $\alpha > 0$, demuéstrese que

$$\alpha(\alpha + 1)\cdots(\alpha + n) = a_n n!\, n^\alpha,$$

donde a_n tiene como cota inferior un número positivo. Demostrar que a_n es monótona decreciente para valores de n suficientemente grandes. [El límite de a_n conforme $n \to \infty$ es $1/\Gamma(\alpha)$.]

3. Encontrar una expresión aproximada para $\dfrac{n_1!\, n_2! \cdots n_l!}{n!}$, donde $n_1 + n_2 + \cdots + n_l = n$.

4. Demostrar que el coeficiente de x^n en el desarrollo binomial de $\dfrac{1}{\sqrt{1 - x}}$ está dado asintóticamente por $\dfrac{1}{\sqrt{\pi n}}$.

7

Sumas y productos infinitos

La serie geométrica, la serie de Taylor y cierto número de ejemplos previamente discutidos en este libro, sugieren la conveniencia de estudiar aquellos procesos de límite pertenecientes al análisis en los cuales la suma de una *serie infinita* aparece desde un punto de vista más general. En principio, cualquier valor límite

$$S = \lim_{n\to\infty} s_n$$

puede expresarse como una serie infinita; sólo se necesita hacer $a_n = s_n - s_{n-1}$ para $n > 1$ y $a_1 = s_1$, obteniéndose

$$s_n = a_1 + a_2 + \cdots + a_n,$$

y así el valor S resulta ser el límite de s_n, la suma de n términos, cuando n crece. Este hecho se expresa diciendo que S es la "suma de la serie infinita"

$$a_1 + a_2 + a_3 + \cdots.$$

Una "suma infinita" tal es simplemente una manera de representar un límite cuando cada aproximación sucesiva se encuentra a partir de la anterior sumando un término más. Así, la expresión decimal de un número a es, en principio, simplemente la representación del número en la forma de una serie infinita $a = a_1 + a_2 + a_3 + \cdots$, en la que, si $0 \leq a \leq 1$, el término a_n se reemplaza por $\alpha_n \times 10^{-n}$, siendo α_n uno de los enteros 0, 1, . . . , 9.

Puesto que todo valor límite puede ser escrito en la forma de una serie infinita, un estudio especial de las series puede parecer superfluo. Sin embargo, tales valores límite aparecen con frecuencia, en forma natural, como series infinitas que exhiben leyes de formación particularmente simples. *No toda* serie tiene una ley de formación fácilmente reconocible. Por ejemplo, el número π puede ciertamente ser representado como un decimal

(que es una serie $\Sigma c_\nu 10^{-\nu}$), aunque no conocemos una ley simple que nos permita establecer el valor de una cifra arbitraria, digamos, la número 7000, de este decimal. Si, no obstante, en vez de ello consideramos la serie de Leibnitz-Gregory para $\pi/4$, tendremos una expresión con una ley general de formación perfectamente clara [ver (7), p. 464].

Análogos a las series infinitas, en las cuales las aproximaciones al límite se forman por la adición repetida de nuevos términos, los llamados *productos infinitos* proporcionan aproximaciones al límite mediante la multiplicación repetida por nuevos factores. Sin embargo, no profundizaremos mucho en la teoría general de los productos infinitos. El tema principal de éste y del capítulo 8 serán las series infinitas.

7.1 Los conceptos de convergencia y divergencia

a. Conceptos básicos

Criterios de convergencia de Cauchy. Considérese una serie infinita con "término general" a_n. La serie [1] es entonces de la forma

$$a_1 + a_2 + \cdots = \sum_{\nu=1}^{n} a_\nu.$$

El símbolo en el lado derecho, en el que aparece el signo de suma, es simplemente una manera abreviada de escribir la expresión de la izquierda.

Si, conforme n crece, la *n-ésima suma parcial*

$$s_n = a_1 + a_2 + \cdots + a_n = \sum_{\nu=1}^{n} a_\nu$$

se aproxima a un límite

$$S = \lim_{n\to\infty} s_n,$$

se dice que la serie es *convergente;* en caso contrario, se dice que es *divergente.* En el primer caso, S se llama la *suma de la serie.*

Hemos encontrado ya muchos casos de series convergentes; como ejemplos podemos citar la serie geométrica $1 + q + q^2 + \cdots$, que converge a la suma $1/(1-q)$ cuando $|q| < 1$, la serie para log 2, la serie para e, y otras.

En el lenguaje de las series infinitas, el criterio de convergencia de Cauchy (ver capítulo 1, pp. 97-98) se expresa como sigue:

---

[1] Para completar el formalismo incluimos la posibilidad de que ciertos a_n se anulen. Si *todos* los términos a partir de cierto índice N (es decir, cuando $n > N$) se anulan, hablamos de una *serie que termina.*

Una condición necesaria y suficiente para la convergencia de una serie es que el número

$$(1) \qquad |s_m - s_n| = |a_{n+1} + a_{n+2} + \cdots + a_m|$$

($m > n$), llegue a ser arbitrariamente pequeño al escogerse para m y n valores suficientemente grandes. En otras palabras: *Una serie es convergente si, y sólo si, se satisface la siguiente condición: dado un número positivo ε, es posible encontrar siempre un índice $N = N(\varepsilon)$ tal que la expresión anterior, $|s_m - s_n|$, sea menor que ε para $m > N$ y $n > N$.*

El criterio de convergencia puede ilustrarse con la serie geométrica para $q = \frac{1}{2}$. Si se toma $\varepsilon = 1/10$, basta hacer $N = 4$. En efecto,

$$|s_m - s_n| = \frac{1}{2^n} + \cdots + \frac{1}{2^{m-1}}$$

$$= \frac{1}{2^{n-1}}\left(\frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{m-n}}\right) < \frac{1}{2^{n-1}}$$

$$\frac{1}{2^{n-1}} < \frac{1}{10} \qquad \text{si } n > 4.$$

Cuando ε se escoge igual a $1/100$, es suficiente con tomar para el correspondiente valor de N el número 7, como puede verificarse fácilmente.

Es obvio que la condición

$$\lim_{n\to\infty} a_n = 0$$

es *necesaria* para la convergencia de una serie. Pues, de lo contrario, el criterio de convergencia daría un resultado negativo para $m = n + 1$. Sin embargo, esta condición necesaria de ninguna manera es *suficiente.* Por el contrario, es fácil encontrar series infinitas cuyos términos generales a_n tienden a 0 al crecer n pero cuyas sumas no existen, debido a que las sumas parciales s_n crecen sin límite al hacerlo n.

Ejemplos. Un ejemplo lo tenemos en la serie

$$1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \cdots + \frac{1}{\sqrt{n}} + \cdots,$$

para la cual el término general es $1/\sqrt{n}$. Se ve de inmediato que

$$s_n > \frac{1}{\sqrt{n}} + \cdots + \frac{1}{\sqrt{n}} = \frac{n}{\sqrt{n}} = \sqrt{n}.$$

La n-ésima suma parcial crece más allá de toda cota al crecer n, por lo que la serie diverge.

Lo mismo puede decirse en el caso, ya clásico, de la *serie armónica*

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots.$$

Aquí

$$a_{n+1} + \cdots + a_{2n} = \frac{1}{n+1} + \cdots + \frac{1}{2n} > \frac{1}{2n} + \cdots + \frac{1}{2n} = \frac{1}{2}.$$

Puesto que n y $m = 2n$ pueden escogerse tan grandes como se desee, la serie diverge, ya que el criterio de Cauchy no es satisfecho. En efecto, por ser positivos todos los términos de la n-ésima suma parcial, ésta tiende a infinito. Por otra parte, la serie formada con los mismos números, pero alternando los signos,

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - + \cdots + \frac{(-1)^{n-1}}{n} + \cdots,$$

converge [ver (4) capítulo 5, p. 462] y tiene por suma log 2.

No hay razón para afirmar que en toda serie divergente S_n tiende a $+\infty$ o a $-\infty$. Así, en la serie infinita

$$1 - 1 + 1 - 1 + 1 + - \cdots,$$

se observa que la suma parcial s_n toma los valores 1 y 0 alternadamente y, como consecuencia de esta oscilación hacia atrás y hacia adelante, ni tiende a un límite definido ni aumenta numéricamente más allá de toda cota.

El hecho siguiente, aunque evidente por sí mismo, es importante y debe hacerse notar. *La convergencia o divergencia de una serie no es alterada insertando o eliminando un número finito de términos.* Por lo que respecta a la convergencia o a la divergencia, carece completamente de importancia el que la serie comience con el término a_0 o con el a_1 o el a_5, o bien cualquier otro término elegido arbitrariamente.

b. Convergencia absoluta y convergencia condicional

La serie armónica $1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} \cdots$, como es sabido, diverge; pero, cuando se cambia en ella el signo de los términos cuyo denominador es par, resulta la serie para log 2, la cual es convergente. Por otra parte, la serie geométrica $1 - q + q^2 - q^3 + - \cdots$, cuando $0 \leq q < 1$, converge y tiene por suma $1/(1+q)$. Haciendo positivos todos los signos que en ella aparecen resulta la serie

$$1 + q + q^2 + q^3 + \cdots,$$

que es también convergente y cuya suma es $1/(1-q)$.

Surge aquí una diferencia que debe ser examinada. Para una serie cuyos términos son todos positivos, existen sólo dos casos posibles: o es convergente o la suma parcial n-ésima crece más allá de toda cota al crecer n. Esto es así porque, constituyendo las sumas parciales una sucesión monótona creciente, convergirían si permaneciesen acotadas. Hay convergencia cuando los términos individuales tienden a cero con rapidez suficiente al crecer n. Se tiene, por otra parte, divergencia si los términos no tienden a cero o bien se acercan a cero con extrema lentitud. Sin embargo, para series con términos positivos y negativos puede ocurrir que los cambios de signo produzcan la convergencia en los casos en que un incremento demasiado grande en las sumas parciales, debido a los términos positivos, sea compensado por los términos negativos, de manera que, como resultado final, dichas sumas tiendan a un límite definido.

Para entender mejor las diferentes posibilidades, considérese una serie $\sum_{\nu=1}^{\infty} a_\nu$ con términos positivos y negativos, y fórmese, para fines de comparación, la serie que tiene los mismos términos pero todos con signos positivos, esto es,

$$|a_1| + |a_2| + \cdots = \sum_{\nu=1}^{\infty} |a_\nu|.$$

Si esta serie converge, entonces, para valores de n y $m > n$ suficientemente grandes, la expresión

$$|a_{n+1}| + |a_{n+2}| + \cdots + |a_m|$$

será, en efecto, tan pequeña como se desee. Teniendo en cuenta que

$$|a_{n+1} + \cdots + a_m| \leq |a_{n+1}| + \cdots + |a_m|$$

la expresión de la izquierda es también arbitrariamente pequeña y, por el criterio de Cauchy, la serie $\sum_{\nu=1}^{\infty} a_\nu$ original converge. En este caso se dice que la serie original es *absolutamente convergente*. Su convergencia se debe a la pequeñez absoluta de sus términos y no depende de cambios en los signos.

Si, por otra parte, la serie con términos a_n diverge en tanto que la serie original converge, se dice que esta última es *condicionalmente convergente*. La convergencia condicional es el resultado de que los términos de signo opuesto se compensan mutuamente.

Criterio de Leibnitz. El *criterio de convergencia de Leibnitz,* enunciado a continuación, es con frecuencia útil para investigar la convergencia condicional:

Si los términos de una serie tienen signos alternados y si, además, sus valores absolutos, $|a_n|$, *tienden monótonamente a* 0 *(de manera que* $|a_{n+1}| \leqslant |a_n|$*), la serie* $\sum_{\nu=1}^{\infty} a_\nu$ *converge.* [Ejemplo: la serie de Leibnitz, (7), p. 464.]

Para la demostración, se supondrá que $a_1 > 0$, lo cual no resta generalidad al razonamiento, y se escribirá la serie en la forma

$$b_1 - b_2 + b_3 - + \cdots,$$

donde todos los términos b_n son ahora positivos, tienden a cero y satisfacen la condición $b_{n+1} \leq b_n$. Agrupando los términos en dos formas diferentes,

$$b_1 - (b_2 - b_3) - (b_4 - b_5) - \cdots$$

y

$$(b_1 - b_2) + (b_3 - b_4) + (b_5 - b_6) + \cdots,$$

se ve de inmediato que las sumas parciales $s_n = \sum_1^n a_\nu$ satisfacen las dos relaciones siguientes

$$s_1 \geq s_3 \geq s_5 \geq \cdots \geq s_{2n+1} \geq \cdots,$$

$$s_2 \leq s_4 \leq s_6 \leq \cdots \leq s_{2n} \leq \cdots.$$

Por otra parte, $s_{2n} \leq s_{2n+1} \leq s_1$ y $s_{2n+1} \geq s_{2n} \geq s_2$. Las sumas parciales impares $s_1, s_3, \ldots$ forman por consiguiente, una sucesión monótona decreciente

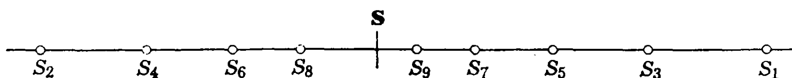


Figura 7.1 Convergencia de una serie alternante.

que nunca cae por debajo de s_2; por tanto, esta sucesión posee un límite, L (p. 96). De la misma manera, las sumas parciales pares $s_2, s_4, \ldots$ forman una sucesión monótona creciente cuyos términos nunca exceden el número fijo s_1, razón por la cual esta sucesión debe tener un valor límite, L'. Puesto que los números s_{2n} y s_{2n+1} difieren entre sí sólo por el número b_{2n+1}, que tiende a cero cuando n crece, los valores límite L y L' son iguales. O sea, las sumas parciales pares e impares tienden al mismo límite, el cual se denotará por S. Esto, sin embargo, implica que la serie es convergente, tal como se afirmó; y su suma es S.

**Criterio de Abel*

Un criterio para la convergencia, que incluye el de Leibnitz como un caso especial, es el *criterio de convergencia de Abel.* Sea $a_1 + a_2 + \cdots$ una serie infinita cuyas sumas parciales $s_n = a_1 + \cdots + a_n$ son acotadas para cualquier valor de n. Sea, además, $p_1, p_2, \ldots$ una sucesión de números positivos que tienden monótonamente a cero. Entonces la serie infinita

$$(2) \qquad p_1 a_1 + p_2 a_2 + \cdots$$

converge. (Para el caso especial de la serie $a_1 + a_2 + \cdots = + 1 - 1 + 1 - 1 + - \cdots$ se encuentra que $p_1 - p_2 + p_3 - \cdots$ converge, lo que no es sino el criterio de

Leibnitz.) La demostración se realiza aplicando el criterio de Cauchy y usando una "suma por partes" para estimar

$$
\begin{aligned}
&|p_{n+1}a_{n+1} + p_{n+2}a_{n+2} + \cdots + p_m a_m| \\
&\quad = |p_{n+1}(s_{n+1} - s_n) + p_{n+2}(s_{n+2} - s_{n+1}) + \cdots + p_m(s_m - s_{m-1})| \\
&\quad = |-p_{n+1}s_n + p_m s_m + (p_{n+1} - p_{n+2})s_{n+1} + (p_{n+2} - p_{n+3})s_{n+2} + \cdots \\
&\qquad + (p_{m-1} - p_m)s_{m-1}| \\
&\quad \leq p_{n+1}M + p_m M + (p_{n+1} - p_{n+2} + p_{n+2} - p_{n+3} + - \cdots + p_{m-1} - p_m)M \\
&\quad = 2p_{n+1}M,
\end{aligned}
$$

donde M es una cota para los valores $|s_i|$. Puesto que $p_{n+1} \to 0$, la convergencia de la serie (2) se desprende del criterio de Cauchy.

*Resulta, en conclusión, otra observación general relativa a la fundamental diferencia que existe entre la convergencia absoluta y la convergencia condicional. Considérese una serie convergente $\sum_{\nu=1}^{\infty} a_\nu$. Denótense sus términos positivos por $p_1, p_2, p_3, \ldots,$ y sus términos negativos por $-q_1, -q_2, -q_3, \ldots$. Al formar la n-ésima suma parcial, $s_n = \sum_{\nu=1}^{\infty} a_\nu$, de la serie dada debe aparecer un cierto número (llamémosle n') de términos positivos y otro cierto número (digamos, n'') de términos negativos, debiendo tenerse $n' + n'' = n$. Además, si el número de términos positivos de la serie, así como también el número de sus términos negativos, es infinito, entonces los números n' y n'' crecerán más allá de toda cota al crecer n. Se ve de inmediato que la suma parcial s_n simplemente es igual a la suma parcial $\sum_{\nu=1}^{n'} p_\nu$, de los términos positivos más la suma parcial, $-\sum_{\nu=1}^{n''} q_\nu$, de los términos negativos. Si la serie dada es absolutamente convergente, la serie de términos positivos, $\sum_{\nu=1}^{\infty} p_\nu$, y la serie de los valores absolutos de los términos negativos, $\sum_{\nu=1}^{\infty} q_\nu$, convergen. Esto es así porque cuando m crece las sumas parciales $\sum_{\nu=1}^{m} p_\nu$ y $\sum_{\nu=1}^{m} q_\nu$ constituyen sucesiones monótonas no decrecientes acotadas superiormente por $\sum_{\nu=1}^{\infty} |a_\nu|$.

La suma de una serie absolutamente convergente es, por lo tanto, igual a la suma de la serie formada por los términos positivos exclusivamente más la suma de la serie que consiste sólo de los términos negativos; en otras palabras, *es igual a la diferencia entre las dos series con términos positivos.*

En efecto, $\sum_{\nu=1}^{n} a_\nu = \sum_{\nu=1}^{n'} p_\nu - \sum_{\nu=1}^{n''} q_\nu$; al crecer n, n' y n'' crecen también

más allá de toda cota, y el límite de la izquierda debe, por tanto, ser igual a la diferencia entre las dos sumas de la derecha. Si la serie contiene solamente un número finito de términos de un signo particular, las cosas, en consecuencia, se simplifican. Si por, otra parte, la serie no converge absolutamente pero lo hace en forma condicional, las serie $\sum_{\nu=1}^{\infty} p_\nu$ y $\sum_{\nu=1}^{\infty} q_\nu$ deben ambas divergir; pues si las dos convergiesen, la serie original debería ser absolutamente convergente, lo que es contrario a la hipótesis. Si sólo una de ellas diverge, digamos, $\sum_{\nu=1}^{\infty} p_\nu$, en tanto que la otra converge, la separación en partes positiva y negativa, $s_n = \sum_{\nu=1}^{n'} p_\nu - \sum_{\nu=1}^{n''} q_\nu$, muestra que la serie no podría ser convergente. En efecto, al crecer n, n' y $\sum_{\nu=1}^{n'} p_\nu$ crecen sin límite, en tanto que el término $\sum_{\nu-1}^{n''} q_\nu$ se aproxima a un límite definido, de manera que la suma parcial s_n crece más allá de toda cota.

Se ve, por tanto, que *una serie condicionalmente convergente no puede considerarse como la diferencia de dos series convergentes, una de ellas formada por sus términos positivos y la otra por los valores absolutos de sus términos negativos.*

En estrecha relación con este hecho, existe otra diferencia entre las series absolutamente convergentes y las series condicionalmente convergentes, de la cual se hace en seguida una breve mención.

*c. Reordenamiento de términos

Las sumas finitas tienen la propiedad de que en ellas es posible cambiar el orden de los términos o, como se dice, reordenar a voluntad los términos, sin que por elo cambie e vaor de lla suma. Surgen en seguida las preguntas: ¿cuál es el significado exacto de un cambio en el orden de los términos de una serie infinita? ¿deja tal reordenamiento inalterado el valor de la suma? Aunque para sumas finitas no existe dificultad en sumar, por ejemplo, los términos en orden inverso, en el caso de una serie infinita tal posibilidad no existe; no hay un último término por el cual empezar. Ahora bien, un cambio de orden en una serie infinita puede sólo significar lo siguiente: se dice que una serie $a_1 + a_2 + a_3 + \cdots$ se transforma mediante un reordenamiento en una serie $b_1 + b_2 + b_3 + \cdots$, si todo término a_n de la primera serie aparece exactamente una vez en la segunda, y recíprocamente. Por ejemplo, a_n puede desplazarse al reordenar, según una cantidad que crezca más allá de toda cota al crecer n. Lo único que debe cumplirse es que a_n aparezca en *alguna parte* de la nueva serie. Si algunos de los términos son desplazados hacia lugares posteriores en la serie, otros términos deberán, por supuesto, ser desplazados hacia sitios anteriores Por ejemplo, la serie

$$1 + q + q^2 + q^4 + q^3 + q^8 + q^7 + q^6 + q^5 + q^{16} + \cdots$$

es un reordenamiento [1] de la serie geométrica $1 + q + q^2 + \cdots$.

[1] Para todo $n > 0$, los términos q^k con $2^n < k \leq 2^{n+1}$ se escriben en orden invertido.

Con respecto al cambio de orden, entre las series absolutamente convergentes y las series condicionalmente convergentes existe una diferencia fundamental:

En las series absolutamente convergentes, un reordenamiento de sus términos no afecta la convergencia y el valor de la suma de la serie queda inalterado, exactamente como en las sumas finitas.

En las series condicionalmente convergentes, por otra parte, el valor de la suma de la serie puede cambiarse a voluntad reordenando adecuadamente sus términos; más aún, por reordenamiento puede hacerse que la serie diverja.

El primero de estos hechos, que se refiere a series absolutamente convergentes, es fácil de establecer. Supóngase inicialmente que la serie en cuestión tiene solamente términos positivos, y considérese la n-ésima suma parcial, $s_n = \sum_{\nu=1}^{n} a_\nu$. Todos los términos de esta suma parcial aparecen en la m-ésima suma parcial $t_m = \sum_{\nu=1}^{m} b_\nu$ de la serie reordenada, con sólo escoger m suficientemente grande. Por tanto $t_m \geq s_n$. Por otra parte, es posible determinar un índice n' tan grande que la suma parcial $s_{n'} = \sum_{\nu=1}^{n'} a_\nu$ de la primera serie contenga todos los términos $b_1, b_2, \ldots, b_m$. Se deduce entonces que $t_m \leq s_{n'} \leq A$, donde A es la suma de la primera serie. Así, para todos los valores de m suficientemente grandes se tiene $s_n \leq t_m \leq A$, y puesto que s_n puede hacerse que difiera de A en una cantidad arbitrariamente pequeña, se concluye que la serie reordenada también converge y, precisamente, lo hace al mismo límite A que la serie original.

Si la serie absolutamente convergente tiene términos tanto positivos como negativos, puede considerarse, en realidad, como la diferencia de dos series formadas exclusivamente por términos positivos. Puesto que al reordenar los términos de la serie original cada una de esas dos series sufre simplemente un reordenamiento y, por consiguiente, converge al mismo valor que antes, la serie original, cuando se reordena, converge también al mismo valor. Pues, por el caso considerado antes, la nueva serie es absolutamente convergente y, por tanto, es la diferencia de las dos series reordenadas consistentes de términos positivos.

Al principiante puede parecer trivial el aserto recién demostrado. Sin embargo, puede mostrarse mediante un ejemplo acerca del comportamiento opuesto de las series condicionalmente convergentes que la demostración es realmente necesaria y que en ella es esencial la convergencia absoluta. Consideremos la ya conocida serie para log 2 y escribamos a continuación el resultado de multiplicarla por el factor $\frac{1}{2}$:

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + - \cdots = \log 2,$$

$$\frac{1}{2} \qquad - \frac{1}{4} \qquad + \frac{1}{6} \qquad - \frac{1}{8} + - \cdots = \frac{1}{2} \log 2 \,;$$

y ahora sumemos, combinando los términos situados en las columnas verticales.[1] Resulta así

$$1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \frac{1}{9} + \frac{1}{11} - \frac{1}{6} + + - \cdots = \frac{3}{2} \log 2.$$

Es claro que esta última serie puede obtenerse reordenando la serie original. Sin embargo, el valor de la suma de la serie ha sido multiplicado por el factor $\frac{3}{2}$. Es fácil imaginar la impresión que debió causar el descubrimiento de esta aparente paradoja a los matemáticos del siglo XVIII, quienes estaban acostumbrados a operar con series infinitas sin considerar cuestiones de convergencia.

---

[1] Para la adición de series véase sección 7.1*d*.

*Damos la demostración del teorema anterior que se refiere al cambio en la suma de una serie condicionalmente convergente, Σa_n, debido a un cambio de orden de sus términos, aunque no tendremos oportunidad de usar tal resultado. Sean $p_1, p_2, \ldots$ los términos positivos y $-q_1, -q_2, \ldots$ los términos negativos de la serie. Como el valor absoluto $|a_n|$ tiende a cero cuando n crece, los números p_n y q_n deben también tender a cero al crecer n. Además, como ya se ha visto, la suma $\sum_1^\infty p_\nu$ debe divergir, así como también $\sum_1^\infty q_\nu$.

Se puede encontrar ahora fácilmente un reordenamiento de la serie original que tenga como suma un número arbitrario a. Supóngase, por conveniencia, que a es positivo. Entonces, súmense los n_1 primeros términos necesarios para justamente lograr que la suma $\sum_1^{n_1} p_\nu$ sea mayor que a. Puesto que la suma $\sum_1^{n_1} p_\nu$ crece con n_1 más allá de toda cota, siempre es posible hacer que la suma parcial sea mayor que a tomando para ello suficientes términos. En estas condiciones, la suma diferirá del valor exacto a en a lo más p_{n_1}. Súmense ahora suficientes términos negativos $-\sum_1^{m_1} q_\nu$ para asegurar que la suma $\sum_1^{n_1} p_\nu - \sum_1^{m_1} q_\nu$ sea menor que a. Esto también es posible debido a la divergencia de la serie $\sum_1^\infty q_\nu$. La diferencia entre esta nueva suma y a es q_{m_1}, a lo sumo. A continuación agréguense otros términos positivos $\sum_{n_1+1}^{n_2} p_\nu$ para hacer que la suma parcial sea de nuevo ligeramente mayor que a, lo cual es posible ya que la serie de términos positivos diverge. La diferencia entre la suma parcial y a es a lo más p_{n_2}. Otra vez, súmense términos negativos $-\sum_{m_1+1}^{m_2} q_\nu$, comenzando con el término siguiente al último usado en el paso anterior, para hacer una vez más que la suma sea menor que a, y prosígase de esta manera. Los valores de las sumas así obtenidas oscilarán en torno al número a; y si el proceso se repite un número suficiente de veces, dichas oscilaciones ocurrirán entre cotas arbitrariamente próximas una de otra. En efecto, la longitud del intervalo en que tales oscilaciones tienen lugar tiende a O ya que los términos p_ν y q_ν mismos tienden a 0 cuando ν crece. En esta forma queda demostrado el teorema.

De manera análoga es posible reordenar la serie para hacerla divergente: sólo se necesita escoger de entre los términos positivos números tan grandes, comparados con los negativos, que la compensación ya no sea posible.

d. Operaciones con series infinitas

Es claro que dos series infinitas que convergen, $a_1 + a_2 + \cdots = S$ y $b_1 + b_2 + \cdots = T$, pueden *sumarse término a término,* esto es, que la serie formada por los términos $c_n = a_n + b_n$ converge y su suma tiene el valor $S + T$.[1] En efecto,

$$\sum_{\nu=1}^{n} c_\nu = \sum_{\nu=1}^{n} a_\nu + \sum_{\nu=1}^{n} b_\nu \to S + T.$$

[1] Este teorema, en realidad, no es sino otra expresión del hecho (ver capítulo 1, p. 95) de que el límite de la suma de dos términos es la suma de sus límites.

También es claro que si se multiplica cada término de una serie infinita, convergente, por el mismo factor, la serie sigue siendo convergente y su suma queda multiplicada por dicho factor.

En estas operaciones no importa si la convergencia es absoluta o condicional. Por otra parte, un estudio más a fondo muestra que la multiplicación de dos series infinitas mediante el método usado para multiplicar sumas finitas no necesariamente da como resultado una serie convergente, a no ser que por lo menos una de las dos series originales sea absolutamente convergente (ver Apéndice, p. 555).

7.2 Criterios de convergencia absoluta y de divergencia

En la sección 7.1*b* se encontró ya el útil criterio de Leibnitz para investigar la convergencia condicional de las series. En las páginas siguientes se considerarán exclusivamente criterios relativos a la convergencia *absoluta.*

a. El criterio de comparación. Mayorantes

Todas las consideraciones de este tipo relativas a la convergencia dependen de la comparación de la serie dada con una segunda serie. Esta se escoge de manera tal que su convergencia pueda investigarse rápidamente. El *criterio general de comparación* puede establecerse como sigue:

Si los números $b_1, b_2, \ldots$ *son todos positivos y la serie* $\sum_{\nu=1}^{\infty} b_\nu$ *converge, y si*

$$|a_n| \leq b_n$$

para todos los valores de n*, entonces la serie* $\sum_{n=1}^{\infty} a_n$ *es absolutamente convergente.*

Usando el criterio de Cauchy, la demostración se vuelve casi trivial. Pues si $m \geq n$, se tiene

$$|a_n + \cdots + a_m| \leq |a_n| + \cdots + |a_m| \leq b_n + \cdots + b_m.$$

Puesto que la serie $\sum_{n=1}^{\infty} b_n$ converge, el miembro de la derecha es arbitrariamente pequeño siempre que n y m sean suficientemente grandes. Como resultado de ello, se tiene que para esos valores de n y m el miembro de la izquierda es también arbitrariamente pequeño, y así, por el criterio de Cauchy, la serie dada converge. La convergencia es absoluta pues nuestro argumento es aplicable también a la convergencia de la serie de los valores absolutos $|a_n|$.

La demostración (análoga) del teorema siguiente puede dejarse al lector. *Si*

$$|a_n| \geq b_n > 0,$$

y la serie $\sum_{n=1}^{\infty} b_n$ *diverge, entonces la serie* $\sum_{n=1}^{\infty} a_n$ *no es absolutamente convergente.*

Algunas veces las series anteriores formadas por los términos positivos b_n se llaman, respectivamente, serie *mayorante* y serie *minorante* de la formada por los términos a_n.

b. Investigación de la convergencia mediante la comparación con la serie geométrica

En las aplicaciones del criterio anterior la serie de comparación más frecuentemente usada como mayorante es la serie geométrica. De inmediato se obtiene el siguiente teorema.

TEOREMA. *La serie* $\sum_{n=1}^{\infty} a_n$ *es absolutamente convergente si a partir de cierto término se cumple una relación de la forma*

$$(3) \qquad |a_n| < cq^n,$$

donde c *es un número positivo independiente de* n *y* q *es cualquier número positivo fijo menor que* 1.

Criterios del cociente y de la raíz. Estos criterios se expresan generalmente en una de las formas siguientes: la serie $\sum_{n=1}^{\infty} a_n$ converge absolutamente si a partir de cierto término se cumple una relación de la forma

$$(4a) \qquad \left|\frac{a_{n+1}}{a_n}\right| < q,$$

donde q es un número positivo menor que 1 e independiente de n, o sea, si a partir de cierto término se cumple una relación de la forma

$$(4b) \qquad \sqrt[n]{|a_n|} < q$$

donde q es un número positivo menor que 1. En particular, las condiciones anteriores se satisfacen si una de las relaciones

$$(5a) \qquad \lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = k < 1$$

o

(5*b*) $$\lim_{n\to\infty} \sqrt[n]{|a_n|} = k < 1$$

es cierta.

Estas afirmaciones se fundamentan fácilmente de la siguiente manera.

Supóngase que (4*a*), el *criterio del cociente,* se satisface a partir de cierto subíndice n_0, esto es, para $n > n_0$. Por brevedad, hágase $a_{n_0+m+1} = b_m$. Se encuentra entonces que

$$|b_1| < q\,|b_0|, \quad |b_2| < q\,|b_1| < q^2\,|b_0|, \quad |b_3| < q\,|b_2| < q^3\,|b_0|,$$

y así sucesivamente. Por tanto,

$$|b_m| < q^m\,|b_0|,$$

y si $n > n_0$ y $c = q^{-n_0-1}\,|b_0|$,

$$\begin{aligned} |a_n| = |b_{n-n_0-1}| &< q^{n-n_0-1}\,|b_0| \\ &= cq^n, \end{aligned}$$

lo que prueba nuestra afirmación. Para el criterio (4*b*), el *criterio de la raíz,* se tiene de inmediato que $|a_n| < q^n$, de lo cual se sigue su veracidad.

Finalmente, para establecer el criterio (5) considérese un número arbitrario, q, tal que $k < q < 1$. Entonces a partir de cierto n_0, esto es, cuando $n > n_0$, las Ecs. 4*a, b* implican que $\left|\dfrac{a_{n+1}}{a_n}\right| < q$ y $\sqrt[n]{|a_n|} < q$, respectivamente, pues a partir de cierto término los valores de $\left|\dfrac{a_{n+1}}{a_n}\right|$ o de $\sqrt[n]{|a_n|}$ difieren de k en menos de $(q - k)$. El criterio queda entonces establecido sobre la base de los resultados demostrados previamente.

Debe hacerse notar que los cuatro criterios, 4*a, b,* 5*a, b,* obtenidos a partir del criterio original $|a_n| < cq^n$, *no* son equivalentes entre sí, ni tampoco equivalentes al original, es decir, no pueden ser derivados unos de otros en ambos sentidos. Pronto veremos, mediante ejemplos, que si una serie satisface una de las condiciones, no necesariamente satisface todas las demás.

Para completar lo anterior, puede señalarse que una serie será, en efecto, divergente si a partir de cierto término se tiene

$$|a_n| > c$$

para algún número positivo c, o bien, si a partir de cierto término

$$\sqrt[n]{|a_n|} > 1,$$

o también si $\displaystyle\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = k,$ o $\displaystyle\lim_{n\to\infty} \sqrt[n]{|a_n|} = k,$

donde k es un número mayor que 1. En efecto, como es fácil ver, en tal serie los términos no pueden tender a cero al crecer n. La serie, por tanto, debe divergir. (En estas condiciones, la serie ni siquiera puede ser condicionalmente convergente.)

Nuestros criterios proporcionan condiciones *suficientes* para la convergencia absoluta de una serie; es decir, cuando se satisfacen puede concluirse que la serie converge absolutamente. Definitivamente, sin embargo, dichas condiciones *no* son *necesarias;* esto es, pueden construirse series absolutamente convergentes que no satisfagan esas condiciones.

Por lo anterior, el saber que

$$\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = 1 \qquad \text{o} \qquad \lim_{n\to\infty}\sqrt[n]{|a_n|} = 1$$

nada implica acerca de la convergencia de la serie; ella puede ser convergente o divergente. Por ejemplo, la serie

$$\sum_{n\to\infty}^{\infty}\frac{1}{n},$$

para la cual $\lim\limits_{n\to\infty}\sqrt[n]{|a_n|} = 1$ y $\lim\limits_{n\to\infty}\left|\dfrac{a_{n+1}}{a_n}\right| = 1$, es divergente, como se vió en la p. 532. Por otra parte, más adelante se verá que la serie $\sum\limits_{n=1}^{\infty}\dfrac{1}{n^2}$; que satisface las mismas condiciones y es convergente.

Como un ejemplo de la aplicación de nuestros criterios, consideremos la serie

$$q + 2q^2 + 3q^3 + \cdots + nq^n + \cdots.$$

Para ella se tiene

$$\lim_{n\to\infty}\sqrt[n]{|a_n|} = |q|\cdot\lim_{n\to\infty}\sqrt[n]{n} = |q|,$$

$$\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = |q|\cdot\lim_{n\to\infty}\frac{n+1}{n} = |q|.$$

La convergencia de la serie cuando $|q| < 1$ se deduce del criterio del cociente y también del criterio de la raíz, aún en la forma más débil dada por (5).

Si, por otra parte, consideramos la serie

$$1 + 2q + q^2 + 2q^3 + \cdots + q^{2n} + 2q^{2n+1} + \cdots,$$

no podremos investigar la convergencia mediante el criterio del cociente cuando $\frac{1}{2} \leq |q| < 1$, pues en ese caso $\left|\dfrac{2q^{2n+1}}{q^{2n}}\right| = 2\,|q| \geq 1$. Sin embargo, el criterio de la raíz inmediatamente dará $\lim\limits_{n\to\infty}\sqrt[n]{|a_n|} = |q|$, mostrando así que la serie converge cuando $|q| < 1$, cosa que, por supuesto, podría también observarse directamente.

c. Comparación con una integral [1]

Procedemos ahora a discutir un método completamente diferente para el estudio de la convergencia. Lo explicaremos para el caso típico, particularmente simple e importante, de la serie

$$\sum_{n=1}^{\infty} \frac{1}{n^\alpha} = 1 + \frac{1}{2^\alpha} + \frac{1}{3^\alpha} + \cdots,$$

en la que el término general, a_n, es $1/n$, siendo α un número positivo. Para investigar la convergencia o divergencia de esta serie, considérese la gráfica de la función $y = 1/x^\alpha$ y márquense sobre el eje x las abscisas enteras $x = 1, x = 2, \ldots$. Para empezar, constrúyase un rectángulo de altura $1/n^\alpha$ sobre el intervalo $n - 1 \leq x \leq n$ del eje x (siendo $n > 1$), y compárese su área con el área de la región limitada por el mismo intervalo, las ordenadas en los extremos y la curva $y = 1/x^\alpha$ (esta región aparece sombreada en la Fig. 7.2). A continuación constrúyase el rectángulo de altura $1/n^\alpha$

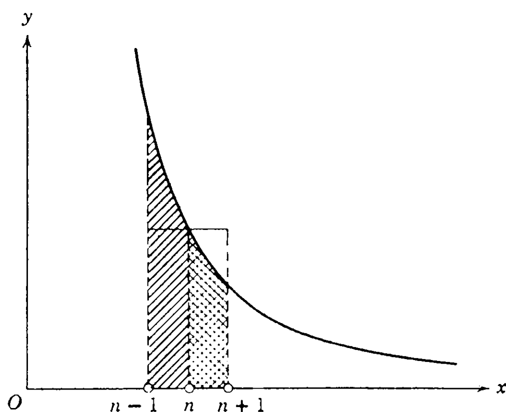


Figura 7.2 Comparación de una serie con una integral.

sobre el intervalo $n \leq x \leq n + 1$, y, de igual forma, compárese con el área de la región que está por encima del intervalo y por debajo de la curva (esta región está indicada en la Fig. 7.2 por el sombreado en cruz). En el primer caso el área bajo la curva es, evidentemente, mayor que el área del rectángulo; en el segundo, es menor. En otras palabras,

$$\int_n^{n+1} \frac{dx}{x^\alpha} < \frac{1}{n^\alpha} < \int_{n-1}^{n} \frac{dx}{x^\alpha}.$$

[1] En relación con esto, véase también el apéndice al capítulo 5, p. 523.

Escribiendo en seguida estas desigualdades para $n = 1, 2, 3, \ldots, m$, y $n = 2, 3, \ldots, m$, respectivamente, y sumando, se obtiene la siguiente estimación para la m-ésima suma parcial $s_m = \sum_{n=1}^{m} \frac{1}{n^\alpha}$:

$$\int_1^{m+1} \frac{dx}{x^\alpha} < s_m < 1 + \int_1^m \frac{dx}{x^\alpha}. \tag{6}$$

Ahora bien, cuando n crece la integral $\int_1^m \frac{1}{x^\alpha} dx$ tiende a un límite finito o bien crece sin límites según que $\alpha > 1$ o $\alpha \leq 1$. En consecuencia, la sucesión monótona de números s_m es acotada o crece más allá de toda cota según que $\alpha > 1$ o $\alpha \leq 1$. En esta forma se tiene el siguiente teorema.

TEOREMA. *La serie de potencias recíprocas,*

$$\sum_{n=1}^{\infty} \frac{1}{n^\alpha} = \frac{1}{1^\alpha} + \frac{1}{2^\alpha} + \frac{1}{3^\alpha} + \cdots,$$

es convergente si, y sólo si, $\alpha > 1$.

Para $\alpha = 1$ resulta, como una consecuencia inmediata, la divergencia de la serie armónica, hecho establecido antes de manera distinta. De manera análoga, las series

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots,$$

$$\frac{1}{1^3} + \frac{1}{2^3} + \frac{1}{3^3} + \cdots,$$

convergen, en tanto que la serie $\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \cdots$ diverge.

La serie convergente $\sum_{\nu=1}^{\infty} \frac{1}{\nu^\alpha}$, con $\alpha > 1$, sirve a menudo como una serie de comparación en la investigación de la convergencia. Por ejemplo, se ve de inmediato que para $\alpha > 1$ la serie $\sum_{\nu=1}^{\infty} \frac{c_\nu}{\nu^\alpha}$ converge absolutamente si los valores absolutos, $|c_\nu|$, de los coeficientes permanecen por debajo de una cota fija independiente de ν.

Constante de Euler. De la estimación (6) con $\alpha = 1$ se deduce en seguida que la sucesión de números $C_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \log n = s_n - \log n > \log(n + 1) - \log n > 0$ está acotada inferiormente. Puesto que, por la desigualdad $\frac{1}{n+1} < \int_n^{n+1} \frac{dx}{x} = \log(n + 1) - \log n = \frac{1}{n+1} +$

$C_n - C_{n+1}$, la sucesión es monótona decreciente, debe tender a un límite:

$$\lim_{n\to\infty} C_n = \lim_{n\to\infty}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \log n\right) = C.$$

El número C, cuyo valor es $0.5772\ldots$, se llama *constante de Euler.* En contraste con otros números especiales e importantes del Análisis, tales como π y e, no se ha encontrado para la constante de Euler otra expresión con una ley de formación sencilla. Hasta la fecha no se sabe si C es racional o irracional.

7.3 Sucesiones de funciones

Como ya se ha hecho resaltar muchas veces, el proceso de límite sirve no sólo para representar números conocidos en términos de otros más simples, sino también para extender el conjunto de números conocidos a otro más amplio. De importancia decisiva en el Análisis es el estudio de los límites, no solamente para sucesiones —o series infinitas— de números constantes, sino, de manera similar, para sucesiones de funciones o series cuyos términos son funciones de una variable x. Este es el caso, por ejemplo, de la serie de Taylor y de las series de potencias en general. Tales procesos de límite se requieren lo mismo para aproximar funciones dadas por otras más simples que para definir y describir analíticamente nuevas funciones, lo que a menudo se realiza sobre la base del concepto de límite para sucesiones de funciones: $f(x) = \lim f_n(x)$ para $n \to \infty$. En forma equivalente, es posible considerar $f(x)$ como la suma y las $f_n(x)$ como las sumas parciales de una serie infinita $f(x) = \sum_{r=1}^{\infty} g_r(x)$ de funciones $g_n(x)$, donde $g_n(x) = f_n(x) - f_{n-1}(x)$ para $n > 1$ y $g_1(x) = f_1(x)$.

Discutiremos a continuación las definiciones precisas y las interpretaciones geométricas.

a. Procesos de límite con funciones y curvas

Definición. La sucesión $f_1(x)$, $f_2(x)$, ... en el intervalo $a \leq x \leq b$ converge a la función límite $f(x)$ si en cada punto x de ese intervalo los valores $f_n(x)$ convergen, en el sentido usual, al valor $f(x)$. En este caso se escribe $\lim_{n\to\infty} f_n(x) = f(x)$. De acuerdo con el criterio de Cauchy (ver p. 98), la convergencia de la sucesión puede expresarse sin hacer referencia a la función límite: la sucesión de funciones converge a una función límite si, y sólo si, en cada punto x del intervalo en cuestión y para cada número positivo ε, la cantidad $|f_n(x) - f_m(x)|$ es menor que ε, siempre que n y m

se tomen suficientemente grandes, es decir, mayores que cierto número. Este número, $N = N(\varepsilon, x)$, depende por lo general de ε y de x y crece más allá de toda cota cuando ε tiende a cero.

Hemos encontrado ya algunos casos de límites de sucesiones de funciones. Mencionamos aquí solamente la definición de la potencia x^α para valores irracionales α mediante la ecuación

$$x^\alpha = \lim_{n \to \infty} x^{r_n},$$

donde $r_1, r_2, \ldots, r_n, \ldots$ es una sucesión de números racionales que tiende a α; y la ecuación

$$e^x = \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n,$$

en la que las funciones de aproximación, $f_n(x)$, en el miembro derecho son polinomios de grado n.

La representación gráfica de funciones por medio de curvas sugiere la posibilidad de hablar de límites de sucesiones de curvas, diciendo, por ejemplo, que las gráficas de las anteriores funciones límite, x^α y e^x, pueden considerarse como las curvas límite de las gráficas de las funciones x^{r_n} y $\left(1 + \frac{x}{n}\right)^n$ respectivamente.

Existe, sin embargo, una diferencia sutil entre el paso al límite con funciones y el paso correspondiente cuando se trata de curvas, la cual no fue vista con claridad sino hasta mediados del siglo XIX. Ilustraremos este punto con un ejemplo y después, en la siguiente sección, lo discutiremos en forma sistemática.

Considérense las funciones

$$f_n(x) = x^n, \qquad n = 1, 2, \ldots,$$

en el intervalo $0 \leq x \leq 1$. Todas estas funciones son continuas y la función límite $\lim_{n \to \infty} f_n(x) = f(x)$ existe. Sin embargo, esta función límite no es continua. Por el contrario, como $f_n(1) = 1$ para todos los valores de n, se tiene el límite

$$f(1) = 1;$$

y, por otra parte, para $0 \leq x < 1$ el límite es $f(x) = \lim_{n \to \infty} f_n(x) = 0$, como se vio en el capítulo 1, p. 88. Por consiguiente, $f(x)$ es una función discontinua que toma el valor 1 en $x = 1$, en tanto que es igual a cero para todo otro valor de x en el intervalo.

Esta discontinuidad queda ilustrada geométricamente por las gráficas C_n de las funciones $y = f_n(x)$. Son éstas (ver Fig. 1.44, p. 89) curvas continuas que pasan por el origen y por el punto $x = 1$, $y = 1$ y que se aproxi-

man más y más al eje x cuando n crece. Las *curvas* tienen una *curva límite* C que no es discontinua sino que consiste (ver Fig. 7.3) de la porción del eje x comprendida entre $x = 0$ y $x = 1$ y la porción de la recta $x = 1$ comprendida entre $y = 0$ e $y = 1$. Las *curvas* convergen, por consiguiente, a una curva límite *continua* que tiene una porción vertical, en tanto que las *funciones* convergen a una función límite *discontinua*. Se ve

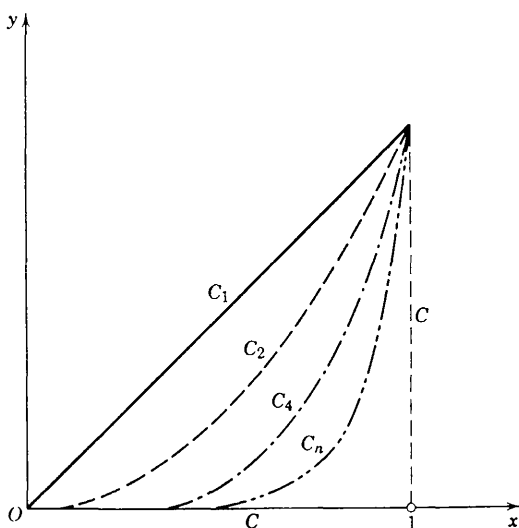


Figura 7.3 Curva límite y función límite.

entonces que esta discontinuidad de la función límite queda expresada por la aparición en la curva límite de una porción perpendicular al eje x. Esta curva límite *no* es la gráfica de la función límite; pues al valor de x para el cual se tiene la porción vertical, la curva hace corresponder un número infinito de valores de y, en tanto que la función le asocia sólo uno. Por consiguiente, el límite de las gráficas de las funciones $f_n(x)$ no es el mismo que la gráfica del límite, $f(x)$, de estas funciones.

7.4 Convergencia uniforme y convergencia no uniforme

a. Comentarios generales y definiciones

La distinción entre los conceptos de convergencia de funciones y de convergencia de curvas es un fenómeno que el estudiante debe comprender con claridad. Involucra la llamada *convergencia no uniforme* de sucesiones o series infinitas de funciones, que será discutida con algún detalle.

Que una función $f(x)$ es el límite de una sucesión $f_1(x)$, $f_2(x)$, ... en un intervalo $a \leq x \leq b$ significa, simplemente por definición, que la relación de límite usual, $f(x) = \lim\limits_{n\to\infty} f_n(x)$, se satisface *en cada punto* del intervalo.

Tal convergencia es una *propiedad local,* en el punto x, de la sucesión. Es natural, sin embargo, pedir algo más que la sola convergencia local de dichas aproximaciones. Se requiere también que si se asigna una medida arbitraria, ε, para la aproximación, entonces, a partir de cierto índice N *todas* las funciones $f_n(x)$ estén comprendidas entre $f(x) - \varepsilon$ y $f(x) + \varepsilon$, para *todos* los valores de x, de manera que sus gráficas $y = f_n(x)$ queden completamente dentro de la franja mostrada en la Fig. 7.4. Se dice que la

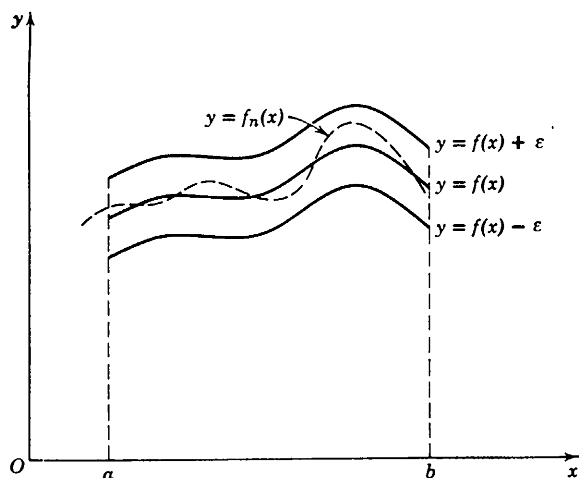


Figura 7.4 Para ilustrar la convergencia uniforme.

aproximación es *uniforme*[1] si su exactitud puede hacerse por lo menos igual a un número positivo prefijado ε, simultáneamente en todos los puntos del intervalo, es decir, escogiendo en todas partes el mismo número $N(\varepsilon)$ independiente de x. Si $\lim\limits_{n\to\infty} f_n(x) = f(x)$ *uniformemente* para $a \leq x \leq b$, existe para todo $\varepsilon > 0$ un correspondiente $N = N(\varepsilon)$ tal que $|f(x) - f_n(x)| < \varepsilon$ para todo $n > N$ y todo x en el intervalo. Mucha gente quedó sorprendida cuando, a mediados del siglo XIX, Seidel y otros observaron que la convergencia de funciones no siempre es uniforme, como ingenuamente se había supuesto.

[1] Compárese con la definición análoga que se dio cuando vimos *continuidad uniforme,* p. 65, según la cual es posible escoger el mismo número $\delta(\varepsilon)$ independientemente de x.

Ejemplos de convergencia no uniforme. El concepto de convergencia uniforme resulta aclarado con algunos ejemplos sobre convergencia no uniforme.

(a) El primer ejemplo lo constituye la sucesión de funciones que se acaba de considerar, $f_n(x) = x^n$. En el intervalo $0 \leq x \leq 1$ esta sucesión converge a la función límite $f(x) = 0$ para $0 \leq x < 1$, $f(1) = 1$. La convergencia se presenta en cada punto del intervalo, es decir, si ε es cualquier número positivo y si se escoge cualquier valor definido y fijo $x = \xi$, entonces la desigualdad $|\xi^n - f(\xi)| < \varepsilon$ se cumple, en efecto, si n es suficientemente grande. Sin embargo, esta aproximación no es uniforme, pues si se toma $\varepsilon = \frac{1}{2}$ puede encontrarse un punto $x = \eta \neq 1$ para el cual $|\eta^n - f(\eta)| = \eta^n > \frac{1}{2}$, sin importar cuán grande se escoja el número n. En efecto, esto es cierto para todos los puntos $x = \eta$ con $1 > \eta > \sqrt[n]{\frac{1}{2}}$. Por consiguiente, es imposible escoger el número n tan grande que la diferencia entre $f(x)$ y $f_n(x)$ sea menor que $\frac{1}{2}$ *en todo el intervalo.*

Este comportamiento se hace comprensible si se observan las gráficas de estas funciones (Fig. 7.3). Por ellas es claro que, no importa cuán grande se escoja el valor de n, para valores de ξ ligeramente menores que 1 el valor de la función $f_n(\xi)$ estará muy próximo a 1 y, por consiguiente, no podrá ser una buena aproximación para $f(\xi)$, que es 0.

Un comportamiento análogo es exhibido por las funciones

$$f_n(x) = \frac{1}{1 + x^{2n}}$$

en la vecindad de los puntos $x = 1$ y $x = -1$. Esto puede ser establecido fácilmente. Aquí $f(x) = 1$ para $|x| < 1$, $f(x) = \frac{1}{2}$ para $x = 1$ y $f(x) = 0$ para $|x| > 1$.

(b) En los dos ejemplos anteriores la no uniformidad de la convergencia tiene relación con el hecho de que *la función límite es discontinua.* Sin embargo, es también fácil construir una sucesión de funciones continuas convergentes a una función límite continua pero sin que la convergencia sea uniforme. Enfoquemos nuestra atención al intervalo $0 \leq x \leq 1$ y hagamos las siguientes definiciones para $n > 2$:

$$f_n(x) = xn^\alpha \qquad \text{para} \qquad 0 \leq x \leq \frac{1}{n},$$

$$f_n(x) = \left(\frac{2}{n} - x\right) n^\alpha \qquad \text{para} \qquad \frac{1}{n} \leq x \leq \frac{2}{n},$$

$$f_n(x) = 0 \qquad \text{para} \qquad \frac{2}{n} \leq x \leq 1,$$

en donde, para comenzar, puede darse a α cualquier valor; pero, una vez hecho esto, debe mantenerse fijo tal valor de α para todos los términos de la sucesión. Gráficamente, nuestras funciones están representadas por una figura en forma de V invertida, consistente en dos segmentos de recta definidos en el intervalo $0 \leq x \leq 2/n$ del eje x, en tanto que de $x = 2/n$ en adelante la gráfica es el eje x mismo (ver Fig. 7.5).

Si $\alpha < 1$, la ordenada del punto más alto de la gráfica, que en general tiene el valor $n^{\alpha-1}$, tenderá a cero al crecer n; y la curva tenderá entonces a confundirse con el eje x, y las funciones $f_n(x)$ convergirán uniformemente a la función límite $f(x) = 0$.

Si $\alpha = 1$, el pico de la gráfica tendrá altura 1 para todo valor de n. Si $\alpha > 1$, la altura del pico crecerá más allá de toda cota al crecer n.

No obstante, sin importar cómo se escoja α, la sucesión $f_1(x)$, $f_2(x)$, ... siempre tiende a la función límite $f(x) = 0$. En efecto, si x es positivo se tiene $2/n < x$ para todos los valores de n suficientemente grandes, de manera que x no está debajo de la parte en forma de V invertida de la gráfica, y $f_n(x) = 0$. Para $x = 0$ todos los valores funcionales $f_n(x)$ son iguales a 0, de manera que en ambos casos

$$\lim_{n\to\infty} f_n(x) = 0.$$

Sin embargo, si $\alpha \geq 1$ la convergencia ciertamente es no uniforme, pues resulta entonces imposible escoger n tan grande que la expresión $|f(x) - f_n(x)| = f_n(x)$ sea menor que $\frac{1}{2}$ en *todo* el intervalo.

(*c*) Un comportamiento exactamente análogo es exhibido por la sucesión de funciones

$$f_n(x) = xn^\alpha e^{-nx},$$

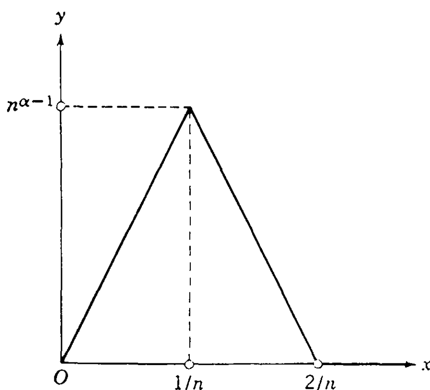


Figura 7.5 Para ilustrar la convergencia no uniforme.

donde, en contraste con el caso anterior, cada función de la sucesión está representada por una única expresión analítica. Aquí, otra vez, la ecuación $\lim_{n\to\infty} f_n(x) = 0$ se cumple para todo valor positivo de x, pues cuando n crece la función e^{-nx} tiende a cero con un orden mayor que el de cualquier potencia de $1/n$ (ver sección 3.7*b*, p. 271). Para $x = 0$ se tiene siempre $f_n(x) = 0$, y así

$$f(x) = \lim_{n\to\infty} f_n(x) = 0$$

para todo valor de x en el intervalo $0 \leq x \leq a$, donde a es un número positivo arbitrario. Sin embargo, aquí la convergencia a la función límite es no uniforme. Pues en el punto $x = 1/n$, en el que $f_n(x)$ alcanza su máximo, se tiene

$$f_n(x) = f_n\left(\frac{1}{n}\right) = \frac{n^{\alpha-1}}{e},$$

de lo cual se sigue que si $\alpha \geq 1$ la convergencia es no uniforme, pues toda curva $y = f_n(x)$, sin importar cuán grande se escoja n, contendrá puntos (a saber, el punto $x = 1/n$, que varía con n) en los cuales $f_n(x) - f(x) = f_n(x) > 1/2e$ (ver Fig. 7.6).

(*d*) Los conceptos de convergencia uniforme y no uniforme pueden, por supuesto, extenderse a una serie infinita. Se dice que una serie

$$g_1(x) + g_2(x) + \cdots$$

es o no uniformemente convergente según sea el comportamiento de sus sumas parciales $f_n(x)$. Un ejemplo muy sencillo de una serie cuya convergencia es no uniforme está dado por

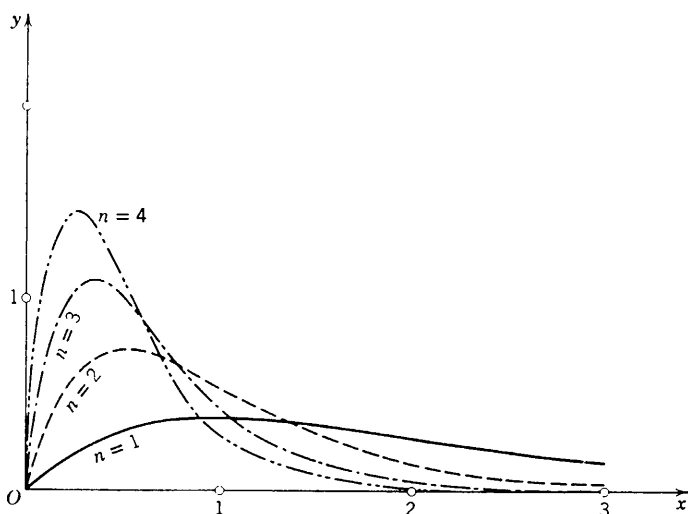


Figura 7.6 Convergencia no uniforme de la sucesión $f_n(x) = n^2xe^{-nx}$.

$$f(x) = x^2 + \frac{x^2}{1 + x^2} + \frac{x^2}{(1 + x^2)^2} + \frac{x^2}{(1 + x^2)^3} + \cdots .$$

Para $x = 0$ toda suma parcial $f_n(x) = x^2 + \cdots + x^2/(1 + x^2)^{n-1}$ tiene el valor 0; y, por tanto, $f(0) = 0$. Para $x \neq 0$ la serie es simplemente una serie geométrica de cociente positivo $1/(1 + x^2) < 1$; y puede, por consiguiente, sumarse usando la: reglas elementales, y así se obtiene para todo $x \neq 0$ la suma

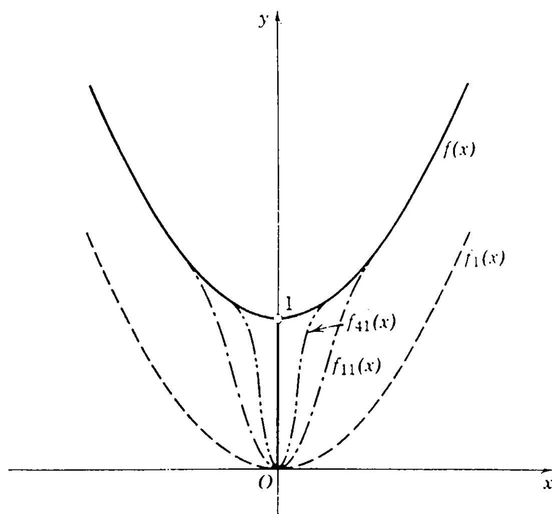


Figura 7.7 Convergencia hacia una función que tiene una discontinuidad por salto eliminable.

$$\frac{x^2}{1 - 1/(1 + x^2)} = 1 + x^2.$$

La función límite $f(x)$ está dada entonces en todas partes excepto en $x = 0$ por la expresión $f(x) = 1 + x^2$, en tanto que $f(0) = 0$. Ella tiene, por consiguiente, una discontinuidad eliminable en el origen.

Una vez más, se tiene aquí convergencia no uniforme en todo intervalo que contiene al origen. Pues la diferencia $f(x) - f_n(x) = r_n(x)$ es siempre igual a 0 para $x = 0$, en tanto que está dada por la expresión $r_n(x) = 1/(1 + x^2)^{n-1}$ para todos los demás valores de x, como puede verificar fácilmente el lector. Si se pide que esta expresión sea menor que, digamos, $\frac{1}{2}$, entonces para cada valor fijo de x esto puede conseguirse escogiendo n suficientemente grande. Pero no puede encontrarse un valor de n lo suficientemente grande para asegurar que $r_n(x)$ sea en todas partes menor que $\frac{1}{2}$; pues si se escoge cualquier valor de n, no importa cuán grande, puede hacerse $r_n(x)$ mayor que $\frac{1}{2}$ con sólo tomar x suficientemente próximo a 0. Por tanto, es imposible una aproximación uniforme con diferencias menores que $\frac{1}{2}$. El asunto se aclara si se consideran las curvas de aproximación (ver Fig. 7.7). Estas curvas, excepto cerca de $x = 0$, quedan cada vez más próximas a la parábola $y = 1 + x^2$ cuando n crece. Cerca de $x = 0$, sin embargo, de las curvas se deriva una protuberancia cada vez más estrecha que se extiende hacia el origen y que, cuando n crece, se aproxima más y más a un cierto segmento de recta, una porción del eje y, de manera que como curva límite se tiene la parábola más una extensión lineal que baja verticalmente hasta el origen.

Como un ejemplo adicional de convergencia no uniforme mencionamos la serie $\sum_{\nu=0}^{\infty} g_\nu(x)$, donde $g_\nu(x) = x^\nu - x^{\nu-1}$ para $\nu \geq 1$, $g_0(x) = 1$, definida en el intervalo $0 \leq x \leq 1$. Las sumas parciales de esta serie son las funciones x^ν ya consideradas en el ejemplo (*a*), p. 549.

b. Un criterio de convergencia uniforme

Las consideraciones anteriores nos muestran que la convergencia uniforme de una sucesión o de una serie es una propiedad especial, que no todas las sucesiones o todas las series tienen. Repetimos ahora la definición de convergencia uniforme según se aplica a las series infinitas: la serie

$$g_1(x) + g_2(x) + \cdots$$

converge uniformemente a una función $f(x)$ en un intervalo si $f(x)$ puede ser aproximada dentro de un margen de aproximación ε (donde ε es un número positivo arbitrariamente pequeño) por la suma de un número suficientemente grande, pero fijo, de términos, $g_1(x) + \cdots + g_N(x) = f_N(x)$, independientemente del punto x del intervalo.

Ahora tenemos un criterio (criterio de Cauchy) para la convergencia uniforme, en el cual no es necesario conocer la función límite $f(x)$: la serie converge uniformemente (o, lo que es lo mismo, la sucesión de funciones $f_n(x)$ converge uniformemente) si, y sólo si, la diferencia $|f_n(x) - f_m(x)|$ puede hacerse, en todo el intervalo, menor que una cantidad positiva arbitraria ε escogiendo n y m mayores que un número N independiente de x. Pues, en primer lugar, si la convergencia es uniforme podemos hacer tanto

$|f_n(x) - f(x)|$ como $|f_m(x) - f(x)|$ menores que $\varepsilon/2$ con sólo escoger n y m mayores que un cierto número N independiente de x; de lo cual se sigue que $|f_n(x) - f_m(x)| < \varepsilon$. Y, en segundo lugar, si $|f_n(x) - f_m(x)| < \varepsilon$ para todo x, y siempre que n y m sean mayores que N, entonces escogiendo cualquier valor fijo de $n > N$ y dejando que m crezca más allá de toda cota se obtiene la relación

$$|f_n(x) - f(x)| = \lim_{m\to\infty} |f_n(x) - f_m(x)| \leq \varepsilon,$$

para todo valor de x, de manera que la convergencia es uniforme.

Como veremos, es precisamente esta condición de convergencia uniforme la que hace que las series infinitas y otros procesos de límite sean herramientas útiles y convenientes en el Análisis. Afortunadamente, en los procesos de límite que, por lo general, se encuentran en el Análisis y sus aplicaciones, la convergencia no uniforme aparece sólo en puntos aislados excepcionales y, por ahora, apenas nos ocasionará algunas dificultades.

Por lo general, la uniformidad de la convergencia de una serie queda establecida mediante el siguiente criterio (comparación de la serie con una serie mayorante de términos constantes):

Si los términos de la serie $\sum_{\nu=1}^{\infty} g_\nu(x)$ *satisfacen la condición* $|g_\nu(x)| \leq a_\nu$, *donde los números* a_ν *son constantes positivas que forman una serie convergente* $\sum_{\nu=1}^{\infty} a_\nu$, entonces la serie $\sum_{\nu=1}^{\infty} g_\nu(x)$ *converge uniformemente (y absolutamente).*

En efecto, se tiene entonces que

$$\left|\sum_{\nu=n}^{m} g_\nu(x)\right| \leq \sum_{\nu=n}^{m} |g_\nu(x)| \leq \sum_{\nu=n}^{m} a_\nu,$$

y puesto que, por el criterio de Cauchy, la suma $\sum_{\nu=n}^{m} a_\nu$ puede hacerse arbitrariamente pequeña con sólo escoger n y $m > n$ suficientemente grandes, esto expresa exactamente la condición necesaria y suficiente para la convergencia uniforme.

Un primer ejemplo lo ofrece la serie geométrica $1 + x + x^2 + \cdots$, con x restringido al intervalo $|x| \leq q$, siendo q cualquier número positivo menor que 1. Los términos de la serie son entonces numéricamente menores que o iguales a los términos de la serie geométrica convergente Σq^ν.

Un segundo ejemplo lo constituye la "serie trigonométrica"

$$\frac{c_1 \operatorname{sen}(x - \delta_1)}{1^2} + \frac{c_2 \operatorname{sen}(x - \delta_2)}{2^2} + \frac{c_3 \operatorname{sen}(x - \delta_3)}{3^2} + \cdots,$$

siempre que $|c_n| < c$, donde c es una constante positiva independiente de n. En efecto, en tal caso se tiene

$$g_n(x) = \frac{c_n \operatorname{sen}(x - \delta_n)}{n^2}, \quad \text{de manera que} \quad |g_n(x)| < \frac{c}{n^2}.$$

Por tanto, la convergencia uniforme y absoluta de la serie trigonométrica se deduce de la convergencia de la serie $\sum_{\nu=1}^{\infty} \frac{c}{n^2}$.

c. Continuidad de la suma de una serie uniformemente convergente de funciones continuas

La importancia de la convergencia uniforme radica en el hecho de que una serie uniformemente convergente se comporta, en muchos aspectos, exactamente como la suma de un número finito de funciones. Así, por ejemplo, la suma de un número finito de funciones continuas es a su vez continua. Conforme a esto, se tiene el siguiente teorema.

TEOREMA. *Si una serie de términos continuos converge uniformemente en un intervalo, su suma es también una función continua.*

DEMOSTRACIÓN. La demostración es muy sencilla. La serie

$$f(x) = g_1(x) + g_2(x) + \cdots$$

se subdivide en la n-ésima suma parcial $f_n(x)$ más el residuo $R_n(x)$. Como ya es costumbre, $f_n(x) = g_1(x) + \cdots + g_n(x)$. Si ahora se asigna cualquier número positivo ε, en virtud de la convergencia uniforme es posible escoger el número n tan grande que el residuo sea menor que $\varepsilon/4$ en todo el intervalo, y, por tanto,

$$|R_n(x + h) - R_n(x)| < \frac{\varepsilon}{2}$$

para cualquier par de números x y $x + h$ del intervalo. La suma parcial $f_n(x)$ consiste en la suma de un número finito de funciones continuas y es, por consiguiente, continua. Por tanto, para cada punto x del intervalo es posible escoger un δ positivo tan pequeño que

$$|f_n(x + h) - f_n(x)| < \frac{\varepsilon}{2}$$

siempre que $|h| < \delta$ y los puntos x y $x + h$ estén en el intervalo. Se deduce entonces que

$$\begin{aligned} |f(x + h) - f(x)| &= |f_n(x + h) - f_n(x) + R_n(x + h) - R_n(x)| \\ &\leq |f_n(x + h) - f_n(x)| + |R_n(x + h) - R_n(x)| < \varepsilon, \end{aligned}$$

lo cual expresa la continuidad de la función.

La importancia de este teorema se pone de relieve cuando recordamos, por nuestros ejemplos anteriores, que las sumas de series convergentes por lo

general no son uniformemente discontinuas. Del teorema anterior podemos concluir que: si la suma de una serie convergente de funciones continuas tiene un punto de discontinuidad, entonces en toda vecindad de este punto la convergencia es no uniforme. Por tanto, toda representación de funciones discontinuas mediante series de funciones continuas debe basarse en el uso de procesos de límite no uniformemente convergentes.

d. Integración de series uniformemente convergentes

La suma de un número finito de funciones continuas puede integrarse "término a término", o sea, la integral de la suma de obtiene integrando cada término separadamente y sumando luego las integrales. En una serie infinita convergente formada por funciones continuas es permisible el mismo procedimiento siempre y cuando la serie converja uniformemente en el intervalo de integración.

Una serie $\sum_{\nu=1}^{\infty} g_\nu(x) = f(x)$ *que converge uniformemente en un intervalo puede integrarse término a término en él.* O bien, en forma más precisa, *si a y x son dos números en el intervalo de convergencia uniforme, la serie* $\sum_{\nu=1}^{\infty} \int_a^x g_\nu(t)\,dt$ *converge y, de hecho, converge uniformemente con respecto a x, siendo su suma igual a* $\int_a^x f(t)\,dt$.[1]

Para demostrar esto, escríbase como antes

$$f(x) = \sum_{\nu=1}^{\infty} g_\nu(x) = f_n(x) + R_n(x).$$

Se ha supuesto que los términos separados de la serie son continuos; y, en consecuencia, por la sección 7.4*c* la suma es también continua y, por tanto, integrable. Ahora bien, si ε es cualquier número positivo, es posible encontrar un número N tan grande que para todo $n > N$ la desigualdad $|R_n(x)| < \varepsilon$ se cumpla para todo valor de x en el intervalo. Por el teorema del valor medio del cálculo integral se tiene

$$\left| \int_a^x [f(t) - f_n(t)]\,dt \right| \leq \varepsilon l,$$

[1] Obsérvese que en este teorema deben tomarse integrales *definidas*. Así, por ejemplo, la serie $\sum_{\nu=1}^{\infty} g_\nu(x)$, con $g_\nu(x) = 0$, converge uniformemente. Sin embargo, tomando la integral indefinida $\int g_\nu(x)\,dx$ = constante = c de cada término, se llega a la serie $\sum_{\nu=1}^{\infty} c$, que, en general, es divergente.

donde l es la longitud del intervalo de integración. Puesto que la integración de la suma finita $f_n(x)$ puede realizarse término a término, resulta

$$\left| \int_a^n f(t)\, dt - \sum_1^x \int_a^x g_\nu(t)\, dt \right| < \varepsilon l.$$

Pero εl puede hacerse tan pequeño como se quiera; por tanto,

$$\sum_{\nu=1}^{\alpha} \int_n^x g_\nu(t)\, dt = \lim_{n\to\infty} \sum_{\nu=1}^{a} \int_a^x g_\nu(t)\, dt = \int_a^x f(t)\, dt,$$

que es lo que debía demostrarse.

Si, en vez de series infinitas, se trata de sucesiones de funciones, el resultado puede expresarse de la manera siguiente:

Si en un intervalo la sucesión de funciones $f_1(x)$, $f_2(x)$, . . . tiende uniformemente a la función límite $f(x)$, entonces

$$\int_a^b f(x)\, dx = \lim_{n\to\infty} \int_a^b f_n(x)\, dx \tag{7}$$

para todo par de números a y b pertenecientes al intervalo. En otras palabras, *pueden intercambiarse las operaciones de integración y de paso al límite.*

Este hecho no es trivial. Desde un punto de vista ingenuo, como el que prevalecía en el siglo XVIII, es cierto que la intercambiabilidad de los dos procesos difícilmente puede ponerse en duda. Sin embargo, un vistazo a los ejemplos de 7.4*a* muestra que en el caso de la convergencia no uniforme la ecuación anterior podría no cumplirse. Basta considerar el ejemplo *b* de p. 549, en el cual la integral de la función límite es 0, en tanto que la integral de la función $f_n(x)$ sobre el intervalo $0 \leq x \leq 1$, es decir, el área del triángulo en la Fig. 7.5, tiene el valor

$$\int_0^1 f_n(x)\, dx = n^{\alpha-2},$$

que no tiende a cero cuando $\alpha \geq 2$. En este caso se ve inmediatamente por la figura que la razón de la diferencia entre $\int_0^1 f(x)\, dx$ y $\lim_{n\to\infty} \int_0^1 f_n(x)\, dx$ radica en la no uniformidad de la convergencia.

Por otra parte, considerando valores de α tales que $1 \leq \alpha < 2$, vemos que la ecuación $\lim_{n\to\infty} \int_0^1 f_n(x)\, dx = \int_0^1 f(x)\, dx$ se satisface aun cuando la convergencia es no uniforme. Como un ejemplo adicional, considérese la serie $\sum_0^\infty g_n(x)$, con $g_n(x) = x^n - x^{n-1}$ para $n \geq 1$ y $g_0(x) = 1$, la que puede integrarse término a término entre los límites 0 y 1 aun cuando no converge uniformemente. Así, aunque la uniformidad de la convergencia es una condición *suficiente* para la integrabilidad término a término, de ninguna manera es una condición *necesaria*.

e. Derivación de series infinitas

El comportamiento de las series o de las sucesiones uniformemente convergentes con respecto a la derivación es completamente diferente de su comportamiento respecto a la integración. Por ejemplo, la sucesión de funciones $f_n(x) = \dfrac{\operatorname{sen} n^2x}{n}$ converge uniformemente a la función límite $f(x) = 0$, pero la derivada $f_n'(x) = n \cos n^2x$ no converge en todas partes a la derivada de la función límite $f'(x) = 0$, como puede verse considerando el punto $x = 0$. Por consiguiente, a pesar de la uniformidad de la convergencia, no es posible intercambiar los procesos de derivación y de paso al límite.

Por supuesto, los correspondientes asertos para series infinitas son válidos. Por ejemplo, la serie

$$\operatorname{sen} x + \frac{\operatorname{sen} 2^4x}{2^2} + \frac{\operatorname{sen} 3^4x}{3^2} + \cdots$$

es absoluta y uniformemente convergente, pues sus términos son numéricamente no mayores que los términos de la serie convergente $\dfrac{1}{1^2} + \dfrac{1}{2^2} + \dfrac{1}{3^2} + \cdots$. Si, no obstante, se deriva la serie término a término, se obtiene la serie

$$\cos x + 2^2 \cos 2^4x + 3^2 \cos 3^4x + \cdots,$$

que es divergente en $x = 0$.

El único criterio útil que asegura en casos especiales que la derivación término a término es permisible queda expresado en el siguiente teorema.

Si al derivar término a término una serie infinita convergente, $\sum_{\nu=0}^{\infty} G_\nu(x) = F(x)$, se obtiene una serie uniformemente convergente de términos continuos, $\sum_{\nu=0}^{\infty} g_\nu(x) = f(x)$, entonces la suma de esta última serie es igual a la derivada de la suma de la primera serie.

Este teorema, por tanto, requiere expresamente que después de derivar término a término la serie se investigue si el resultado de tal derivación es o no una serie uniformemente convergente.

La demostración del teorema es casi trivial. En efecto, por el teorema de la sección 7.4*d*, la serie obtenida por derivación puede integrarse término a término. Teniendo en cuenta que $g_\nu(t) = G_\nu'(t)$, se obtiene

$$\int_a^x f(t)\,dt = \int_a^x \left(\sum_0^\infty g_\nu(t)\right) dt = \sum_{\nu=0}^\infty \int_a^x g_\nu(t)\,dt = \sum_{\nu=0}^\infty (G_\nu(x) - G_\nu(a))$$
$$= F(x) - F(a).$$

Lo cual, siendo cierto para todo valor de x en el intervalo de convergencia uniforme, implica que

$$f(x) = F'(x),$$

como se quería demostrar.

7.5 Series de potencias

Las *series de potencias* ocupan un lugar preponderante entre las series infinitas. Una serie de potencias es una serie del tipo

$$(8) \qquad P(x) = c_0 + c_1x + c_2x^2 + \cdots = \sum_{\nu=0}^{\infty} c_\nu x^\nu$$

("serie de potencias en x"), o, más generalmente,

$$(8a) \quad P(x) = c_0 + c_1(x - x_0) + c_2(x - x_0)^2 + \cdots = \sum_{\nu=0}^{\infty} c_\nu(x - x_0)^\nu$$

("serie de potencias en $(x - x_0)$"), donde x_0 es un número fijo. Si en esta última serie introducimos $\xi = x - x_0$ como una nueva variable, obtenemos una serie de potencias $\sum_{\nu=0}^{\infty} c_\nu \xi^\nu$ en la nueva variable ξ, por lo que podemos enfocar nuestra atención a las series de potencias de la forma especial $\sum_{\nu=0}^{\infty} c_\nu x^\nu$ sin pérdida de generalidad.

En el capítulo 5 (p. 465) consideramos la representación aproximada de funciones mediante polinomios, y en esa forma llegamos al desarrollo de funciones en serie de Taylor, las cuales son, de hecho, series de potencias. En esta sección estudiaremos con más detalle las series de potencias y obtendremos, en una forma más conveniente que antes, los desarrollos en series de algunas de las funciones más notables.

a. Propiedades de convergencia de las series de potencias—Intervalo de convergencia

Existen series de potencias que para ningún valor de x convergen, excepto, claro está, para $x = 0$. Como ejemplo tenemos la serie

$$x + 2^2x^2 + 3^3x^3 + \cdots + n^nx^n + \cdots.$$

En efecto, si $x \neq 0$ podemos encontrar un entero N tal que $|x| > 1/N$. Entonces, todos los términos n^nx^n para los cuales $n > N$ serán mayores que 1 en valor absoluto y, de hecho, cuando n crece n^nx^n crece más allá de toda cota, por lo que la serie no puede ser convergente.

Por otra parte, existen series que convergen para *todo* valor de x; por ejemplo, la serie de potencias para la función exponencial

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots,$$

cuya convergencia para todo valor de x se deduce de inmediato a partir del criterio del cociente (criterio 5*a*, p. 540). En $(n+1)$-ésimo término dividido por el n-ésimo término da x/n, y, cualquiera que sea el número x escogido, este cociente tiende a cero al crecer n.

El comportamiento de las series de potencias respecto a la convergencia es expresado en el siguiente teorema fundamental.

Si una serie de potencias en x converge para un valor $x = \xi$, converge absolutamente para todo valor de x tal que $|x| < |\xi|$, y la convergencia es uniforme en todo intervalo $|x| \leq \eta$ siendo η cualquier número positivo menor que $|\xi|$. Aquí η puede estar tan próximo a $|\xi|$ como se quiera.

La demostración es sencilla. Si la serie $\sum_{\nu=0}^{\infty} c_\nu \xi^\nu$ converge, sus términos tienden a cero cuando n crece. De aquí se desprende el aserto, más débil, de que todos los términos están por debajo de una cota M independiente de ν, es decir, $|c_\nu \xi^\nu| < M$. Si en estas condiciones representamos por q un número cualquiera que satisfaga la condición $0 < q < 1$, y si x se restringe al intervalo $|x| \leq q\,|\xi|$, entonces $|c_\nu x^\nu| \leq |c_\nu \xi^\nu|\, q^\nu < Mq^\nu$. En este intervalo, por consiguiente, los términos de la serie $\sum_{0}^{\infty} c_\nu x^\nu$ son en valor absoluto más pequeños que los términos de la serie geométrica convergente $\Sigma M q^\nu$. Por tanto, del teorema de la p. 553 se deduce la convergencia absoluta y uniforme de la serie en el intervalo $-q\,|\xi| \leq x \leq q\,|\xi|$.

Si una serie de potencias no converge en todas partes, es decir, si existe un valor $x = \xi$ para el cual la serie diverge, debe divergir para todo valor de x tal que $|x| > |\xi|$. Pues si fuese convergente para tal valor de x, por el teorema anterior tendría que convergir para el valor ξ numéricamente menor.

De lo anterior se desprende que una serie de potencias que converge al menos para un valor de x diferente de 0 y que diverge al menos para un valor de x, tiene un *intervalo de convergencia;* esto es, existe un número positivo determinado ρ tal que para $|x| > \rho$ la serie diverge y para $|x| < \rho$ la serie converge. Para $|x| = \rho$ no es posible hacer afirmación general alguna. Aquí ρ es precisamente el *extremo superior* de los valores de x para los cuales la serie converge (y la existencia de tal extremo superior queda garantizada por el teorema de la p. 121, pues los valores de x para los cuales la serie converge forman un conjunto acotado). Los casos límite,

aquéllos en los que la serie converge solamente para $x = 0$ y aquéllos en que converge en todas partes, son expresados simbólicamente escribiendo $\rho = 0$ y $\rho = \infty$ respectivamente.[1]

Por ejemplo, para la serie geométrica $1 + x + x^2 + \cdots$ tenemos $\rho = 1$. En los puntos extremos del intervalo de convergencia la serie diverge. De manera análoga, para la serie de la tangente inversa (p. 463),

$$\operatorname{arc\,tan} x = x - \frac{x^3}{3} + \frac{x^5}{5} - + \cdots,$$

tenemos $\rho = 1$, y en ambos puntos extremos ($x = \pm 1$) del intervalo de convergencia la serie converge, lo cual puede verse usando el criterio de Leibnitz (p. 533).

De la convergencia uniforme se deriva el importante resultado de que dentro de su intervalo de convergencia (si tal intervalo existe) una serie de potencias representa una función continua.

b. Integración y derivación de series de potencias

Por la uniformidad de la convergencia, *siempre es válido integrar término a término una serie de potencias*

$$f(x) = \sum_{\nu=0}^{\infty} c_\nu x^\nu$$

sobre cualquier intervalo cerrado contenido completamente por el intervalo de convergencia. Obtenemos así la función

$$F(x) = c + \sum_{\nu=0}^{\infty} \frac{c_\nu}{\nu + 1} x^{\nu+1}, \tag{9}$$

para la cual

$$F'(x) = f(x) \quad \text{y} \quad F(0) = c.$$

También es posible derivar término a término una serie de potencias dentro de su intervalo de convergencia, obteniéndose así la ecuación

$$f'(x) = \sum_{\nu=1}^{\infty} \nu c_\nu x^{\nu-1}. \tag{10}$$

Para demostrar esta afirmación basta probar que la serie de la derecha converge uniformemente si x se restringe a un intervalo contenido com-

[1] Es posible encontrar este intervalo de convergencia directamente a partir de los coeficientes c_ν de la serie. Si el límite $\lim_{n\to\infty} \sqrt[n]{|c_n|}$ existe, entonces

$$\rho = \frac{1}{\lim_{n\to\infty} \sqrt[n]{|c_n|}}.$$

Para el caso general véase problema 8, p. 586.

pletamente por el intervalo de convergencia. Por tanto, supóngase que ξ es un número, tan próximo a ρ como se quiera, para el cual $\sum_{\nu=1}^{\infty} c_\nu \xi^\nu$ converge. Entonces, como ya se ha visto, los números $|c_\nu \xi^\nu|$ quedan todos por debajo de una cota M independiente de ν, de manera que $|c_\nu \xi^{\nu-1}| < \frac{M}{|\xi|} = N$. Ahora bien, sea q cualquier número tal que $0 < q < 1$. Si x se restringe al intervalo $|x| \leq q\,|\xi|$, los términos de la serie infinita (10) no son mayores que los de la serie $\sum_{\nu=1}^{\infty} |\nu c_\nu \, q^{\nu-1} \, \xi^{\nu-1}|$, y, por consiguiente, son menores que los términos de la serie $\sum_{\nu=0}^{\infty} N\nu q^{\nu-1}\cdot$ Sin embargo, en esta última serie la razón del $(n+1)$-ésimo término sobre el n-ésimo es $q(n+1)/n$, que tiende a q cuando n crece. Como $0 < q < 1$, se concluye [criterio (5a)] que esta serie converge. Por tanto, la serie obtenida por derivación converge uniformemente y, por el teorema de la p. 557, representa la derivada $f'(x)$ de la función $f(x)$, lo que prueba nuestra afirmación.

Si este resultado se aplica nuevamente a la serie de potencias

$$f'(x) = \sum_{\nu=1}^{\infty} \nu c_\nu x^{\nu-1},$$

al derivar término a término se encuentra que

$$f''(x) = \sum_{\nu=2}^{\infty} \nu(\nu-1) c_\nu x^{\nu-2}.$$

y, continuando el proceso, se llega al teorema: *Toda función representada por una serie de potencias puede derivarse tantas veces como se desee dentro del intervalo de convergencia, y esta derivación puede realizarse término a término.*[1]

c. Operaciones con series de potencias

Los teoremas anteriores relativos al comportamiento de las series de potencias son la justificación para operar con series de potencias de la misma manera que con polinomios. Es obvio que dos series de potencias pueden

[1] Como una expresión explícita para la k-ésima derivada obtenemos

$$f^{(k)}(x) = \sum_{\nu=k}^{\infty} \nu(\nu-1)\cdots(\nu-k+1) c\, x^{\nu-k},$$

o, en forma ligeramente diferente,

$$\frac{f^{(k)}(x)}{k!} = \sum_{\nu=k}^{\infty} \binom{\nu}{k} c_\nu x^{\nu-k} = \sum_{\nu=0}^{\infty} \binom{k+\nu}{k} c_{k+\nu} x^\nu.$$

Estas dos fórmulas son a menudo útiles.

sumarse o restarse sumando o restando los coeficientes correspondientes (ver p. 538). También es claro que una serie de potencias, lo mismo que cualquier otra serie convergente, puede multiplicarse por un factor constante multiplicando cada término por ese factor. Por otra parte, la multiplicación y la división de dos series requieren un estudio algo más detallado, para el cual remitimos al lector al Apéndice (p. 572). Aquí mencionamos solamente, sin demostrarlo, que dos series de potencias,

$$f(x) = \sum_{\nu=0}^{\infty} a_\nu x^\nu$$

y

$$g(x) = \sum_{\nu=0}^{\infty} b_\nu x^\nu,$$

pueden multiplicarse entre sí, como los polinomios. Concretamente, tenemos los siguientes teoremas: *En toda la parte común de sus intervalos de convergencia el producto de esas dos series está dado por la serie de potencias convergente* $\sum_{\nu=0}^{\infty} c_\nu x^\nu$, *en la que los coeficientes están dados por las fórmulas*

$$\begin{aligned} c_0 &= a_0 b_0, \\ c_1 &= a_0 b_1 + a_1 b_0, \\ c_2 &= a_0 b_2 + a_1 b_1 + a_2 b_0, \\ &\cdots\cdots\cdots\cdots \\ c_n &= a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0, \\ &\cdots\cdots\cdots\cdots \end{aligned}$$

d. Unicidad del desarrollo

En la teoría de las series de potencias es importante el hecho siguiente: si dos series de potencias, $\sum_{\nu=0}^{\infty} a_\nu x^\nu$ y $\sum_{\nu=0}^{\infty} b_\nu x^\nu$, convergen en un intervalo que contiene al punto $x = 0$ en su interior, y si en ese intervalo las dos series representan la misma función $f(x)$, entonces son idénticas entre sí, es decir, la ecuación $a_n = b_n$ es válida para todo valor de n. En otras palabras:

Una función $f(x)$ puede representarse mediante una serie de potencias en x de una única manera.

Brevemente: la representación de una función mediante una serie de potencias es "única".

Para la demostración sólo se necesita observar que la diferencia de dos series de potencias, esto es, la serie de potencias $\phi(x) = \sum_{\nu=0}^{\infty} c_\nu x^\nu$ con coeficientes $c_\nu = a_\nu - b_\nu$, representa la función

$$\phi(x) = f(x) - f(x) = 0$$

en el intervalo; es decir, esta última serie de potencias converge al límite 0 en todo el intervalo. En particular, para $x = 0$ la suma de la serie debe ser 0; o sea, $c_0 = 0$, de manera que $a_0 = b_0$. Ahora derivamos la serie en el interior del intervalo; y así resulta $\phi'(x) = \sum_{\nu=1}^{\infty} \nu c_\nu x^{\nu-1}$. Sin embargo, $\phi'(x)$ es también 0 en todo el intervalo; por tanto, para el caso particular de $x = 0$ se tiene $c_1 = 0$, o sea, $a_1 = b_1$. Continuando con este proceso consistente en derivar y luego hacer $x = 0$, se encuentra sucesivamente que todos los coeficientes c_ν son iguales a cero, lo que prueba nuestro teorema.

Además, de esta discusión puede extraerse la conclusión siguiente: si se toma la ν-ésima derivada de una serie $f(x) = \Sigma a_\nu x^\nu$ y luego se hace $x = 0$, se obtiene de inmediato

$$a_\nu = \frac{1}{\nu!} f^{(\nu)}(0),$$

esto es:

Toda serie de potencias que converge en puntos distintos de $x = 0$ es la serie de Taylor de la función que representa.

La unicidad del desarrollo corresponde al hecho de que los coeficientes pueden ser expresados en términos de la función misma.

*e. Funciones analíticas

El nombre "funciones analíticas" ha sido usado para designar las funciones $f(x)$ que pueden ser expresadas mediante series de potencias, desde que la importancia de tales funciones fue rconocida por Lagrange. Específicamente, $f(x)$ se llama ***analítica*** en la vecindad de $x = a$ si en esta vecindad es posible desarrollar $f(x)$ como una serie de potencias en $x - a$ que sea convergente.

En tanto que las funciones que no son totalmente o en todas partes analíticas desempeñan un papel central en el Análisis y sus aplicaciones, las funciones analíticas son particularmente importantes porque comparten con los polinomios muchas características simples.

Por ejemplo, una función analítica que no es idénticamente nula tendrá alguna derivada no nula en $x = a$. Sea r el número más pequeño para el cual $f^{(r)}(a) \neq 0$. Entonces, teniendo f un cero de orden r en el punto $x = a$, puede ser representada como un producto $f(x) = (x - a)^r g(x)$, donde $g(x)$ es una función analítica para la cual $g(a) = \frac{1}{r!} f^{(r)}(a)$ es diferente de cero. (Compárese con capítulo 5, p. 482.) Sin duda, la posibilidad de sacar como factor común la potencia $(x - a)^r$ se deduce de inmediato de la convergencia de la serie de potencias respectiva.

También, como se desprende de la continuidad de la serie de potencias convergente que se tiene para $g(x)$, el factor $g(x)$ no se anula en una vecindad adecuada de $x = a$, o sea: *los ceros de $f(x)$ están aislados,* a menos, claro está, que f se anule idénticamente.

Como lo mismo es cierto para la función $f'(x)$, se concluye que en un intervalo finito una función analítica es monótona por pedazos, es decir, no puede cambiar su carácter de monotonía un número infinito de veces. Así, la gráfica de $y = f(x)$ no puede intersectar en una infinidad de puntos a una recta $y =$ constante (o cualquier línea recta) dentro de un intervalo finito.

Puede observarse que estas últimas afirmaciones no son ciertas necesariamente para funciones no analíticas, como ocurre para $y = \operatorname{sen}(1/x)\, e^{-1/x^2}$ en la vecindad de $x = 0$ (ver p. 481).

7.6 Desarrollos en series de potencias de funciones dadas. El método de los coeficientes indeterminados. Ejemplos

Toda serie de potencias representa dentro de su intervalo de convergencia una función continua con derivadas de todos los órdenes. Discutiremos ahora el problema inverso consistente en desarrollar en una serie de potencias una función dada. Teóricamente esto siempre puede hacerse por medio del teorema de Taylor. En la práctica surgen a menudo dificultades en el cálculo concreto de la n-ésima derivada y en la estimación del residuo. Sin embargo, en muchos casos es posible alcanzar dicho objetivo haciendo uso del siguiente artificio. Como primer paso se escribe $f(x) = \sum_{\nu=0}^{\infty} c_\nu x^\nu$, donde los coeficientes c_ν son inicialmente desconocidos. A continuación, usando alguna propiedad conocida de la función $f(x)$ se determinan los coeficientes, y luego se demuestra que la serie converge. Esta serie representa una función, y lo único que falta es demostrar que tal función es idéntica a $f(x)$. Por la unicidad del desarrollo en serie de potencias, se sabe que el desarrollo deseado lo constituye la serie obtenida, y ninguna otra. En realidad, las series para $\arctan x$ y $\log(1 + x)$ fueron obtenidas ya por un método relacionado con la idea que se expone en este capítulo; pues simplemente se integraron término a término las series para las derivadas de estas funciones, las cuales se sabía eran series geométricas. Consideraremos ahora algunos ejemplos de este método.

a. La función exponencial

Como se vio en el capítulo 3, sección 4*a*, p. 244, la función $y = e^x$ está completamente caracterizada por la ecuación diferencial $y' = y$ y la condición inicial $y = 1$ para $x = 0$. Estas propiedades pueden usarse directamente para encontrar la serie de potencias de la función exponencial. El problema consiste en hallar una función $f(x)$ para la cual $f'(x) = f(x)$ y $f(0) = 1$. Si se escribe formalmente la serie con sus coeficientes indeterminados,

$$f(x) = c_0 + c_1 x + c_2 x^2 + \cdots.$$

y se la deriva, se obtiene

$$f'(x) = c_1 + 2c_2 x + 3c_3 x^2 + \cdots.$$

Puesto que, por hipótesis, estas dos series de potencias deben ser idénticas, resulta la ecuación

$$nc_n = c_{n-1},$$

válida para todos los valores de $n \geq 1$. Si se observa que, por la relación $f(0) = 1$, el coeficiente c_0 debe tener el valor 1, es posible calcular entonces todos los coeficientes en forma sucesiva y de esta manera obtener la serie de potencias

$$f(x) = 1 + \frac{x}{1!} + \frac{x^3}{2!} + \frac{x^3}{3!} + \cdots.$$

Como se deduce fácilmente del criterio del cociente, esta serie converge para cualquier valor de x y, por consiguiente, representa una función para la cual las relaciones $f'(x) = f(x)$ y $f(0) = 1$ efectivamente se cumplen. (Intencionalmente se ha evitado aquí cualquier uso de lo previamente aprendido acerca del desarrollo de la función exponencial.)

Como únicamente la función e^x posee estas propiedades, se concluye en seguida que la función $f(x)$ es idéntica a e^x.

b. La serie del binomio

Podemos ahora volver a la serie del binomio (sección 5.5*c*, p. 475), haciendo uso esta vez del método de los coeficientes indeterminados. Deseamos desarrollar la función $f(x) = (1 + x)$ en una serie de potencias, por lo cual escribimos

$$f(x) = (1 + x)^\alpha = c_0 + c_1 x + c_2 x^2 + \cdots,$$

estando indeterminados los coeficientes c_ν. Observamos ahora que la función satisface obviamente la relación

$$(1 + x)f'(x) = \alpha f(x) = \sum_{\nu=0}^{\infty} \alpha c_\nu x^\nu:$$

Por otra parte, si derivamos término a término la serie para $f(x)$ y el resultado lo multiplicamos por $(1 + x)$, obtenemos

$$(1 + x)f'(x) = c_1 + (2c_2 + c_1)x + (3c_3 + 2c_2)x^2 + \cdots.$$

Ahora bien, puesto que estas dos series de potencias para $(1 + x)f'(x)$ deben ser idénticas entre sí, debe ser

$$\alpha c_0 = c_1, \quad \alpha c_1 = 2c_2 + c_1, \quad \alpha c_2 = 3c_3 + 2c_2, \ldots.$$

Además, como nuestra serie debe tener el valor 1 para $x = 0$, se encuentra que $c_0 = 1$, y así resultan sucesivamente las expresiones

$$c_1 = \alpha, \quad c_2 = \frac{(\alpha - 1)\alpha}{2}, \quad c_3 = \frac{(\alpha - 2)(\alpha - 1)\alpha}{3 \cdot 2}, \ldots,$$

para los coeficientes. En general se tiene

$$c_\nu = \frac{(\alpha - \nu + 1)(\alpha - \nu + 2)\cdots(\alpha - 1)\alpha}{\nu(\nu - 1)\cdots 2 \cdot 1} = \binom{\alpha}{\nu}.$$

como puede establecerse fácilmente. Al substituir estos valores para los coeficientes, resulta la serie $\sum_{\nu=0}^{\infty} \binom{\alpha}{\nu} x^\nu$. Hasta aquí no se ha investigado la convergencia de la serie ni demostrado que ella representa realmente la función $(1 + x)$.

Por el criterio del cociente, se encuentra que cuando α es un entero no positivo la serie converge si $|x| < 1$ y diverge si $|x| > 1$; pues en ese caso la relación entre los términos $(n + 1)$-ésimo y n-ésimo es $\dfrac{\alpha - n + 1}{n} x$, y el valor absoluto de esta expresión tiende a x cuando n crece más allá de toda cota.[1] Por tanto, si $|x| < 1$ la serie representa una función $f(x)$ que satisface la condición $(1 + x)f'(x) = \alpha f(x)$, según

[1] Mencionamos aquí, sin demostración, las condiciones exactas bajo las cuales esta serie converge. Si el índice α es un entero ≥ 0, la serie termina, y es válida, por consiguiente, para todos los valores de x (reduciéndose al teorema del binomio ordinario). Para todo otro valor de α la serie es absolutamente convergente si $|x| < 1$ y divergente si $|x| > 1$. Para $x = +1$ la serie converge absolutamente si $\alpha > 0$, converge condicionalmente si $-1 < \alpha < 0$ y diverge si $\alpha \leq -1$. Finalmente, para $x = -1$ la serie es absolutamente convergente si $\alpha > 0$, y es divergente si $\alpha < 0$.

se deduce del método usado para determinar los coeficientes. Además, $f(0) = 1$. Estas dos condiciones juntas garantizan que la función $f(x)$ es idéntica a $(1 + x)^\alpha$. En efecto, haciendo

$$\phi(x) = \frac{f(x)}{(1 + x)^\alpha}$$

se encuentra que

$$\phi'(x) = \frac{(1 + x)^\alpha f'(x) - \alpha(1 + x)^{\alpha-1}f(x)}{(1 + x)^{2\alpha}} = 0;$$

por tanto, $\phi(x)$ es una constante y, de hecho, es siempre igual a 1 ya que $\phi(0) = 1$. Hemos, pues, probado que para $|x| < 1$ se tiene

$$(1 + x)^\alpha = \sum_{\nu=0}^{\infty} \binom{\alpha}{\nu} x^\nu,$$

que es la serie del binomio.

Mencionamos aquí los siguientes casos especiales de la serie del binobio: la serie geométrica

$$\frac{1}{1 + x} = (1 + x)^{-1} = 1 - x + x^2 - x^3 + x^4 - + \cdots$$

$$= \sum_{\nu=0}^{\infty} (-1)^\nu x^\nu$$

la serie

$$\frac{1}{(1 + x)^2} = (1 + x)^{-2} = 1 - 2x + 3x^2 - 4x^3 + - \cdots$$

$$= \sum_{\nu=0}^{\infty} (-1)^\nu (\nu + 1) x^\nu;$$

que también puede obtenerse por derivación a partir de la serie geométrica; y la serie

$$\sqrt{(1 + x)} = (1 + x)^{1/2} = 1 + \frac{1}{2}x - \frac{1}{2 \cdot 4}x^2 + \frac{1 \cdot 3}{2 \cdot 4 \cdot 6}x^3$$

$$- \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6 \cdot 8}x^4 + - \cdots,$$

$$\frac{1}{\sqrt{(1 + x)}} = (1 + x)^{-1/2} = 1 - \frac{1}{2}x + \frac{1 \cdot 3}{2 \cdot 4}x^2 - \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6}x^3$$

$$+ \frac{1 \cdot 3 \cdot 5 \cdot 7}{2 \cdot 4 \cdot 6 \cdot 8}x^4 - + \cdots,$$

para la cual constituyen buenas aproximaciones los dos o tres primeros términos.

c. La serie para arc sen x

Tal serie puede obtenerse de manera muy sencilla desarrollando la expresión $1/\sqrt{(1 - t^2)}$ de acuerdo con la serie del binomio:

$$(1 - t^2)^{-1/2} = 1 + \frac{1}{2}t^2 + \frac{1 \cdot 3}{2 \cdot 4}t^4 + \cdots.$$

Esta serie converge si $|t| \leq 1$, de manera que converge uniformemente si $|t| \leq q < 1$. Al integrar término a término entre 0 y x, resulta

$$\text{arc sen}\, x = x + \frac{1}{2}\frac{x^3}{3} + \frac{1 \cdot 3}{2 \cdot 4}\frac{x^5}{5} + \cdots,$$

la cual, por el criterio del cociente, converge si $|x| < 1$ y diverge si $|x| > 1$.

La obtención de esta serie a partir del teorema de Taylor sería menos conveniente, debido a la dificultad en estimar el residuo.

d. La serie para arc senh $x = \log [x + \sqrt{(1 + x^2)}]$

Este desarrollo se obtiene de manera análoga. Usando el anterior teorema de la serie del binomio, escribimos la serie para la derivada de arc senh x.

$$\frac{1}{\sqrt{1 + x^2}} = 1 - \frac{1}{2}x^2 + \frac{1 \cdot 3}{2 \cdot 4}x^4 - \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6}x^6 + - \cdots,$$

y luego integramos término a término. Así se obtiene el desarrollo

$$\text{ar senh}\, x = x - \frac{1}{2}\frac{x^3}{3} + \frac{1 \cdot 3}{2 \cdot 4}\frac{x^5}{5} - + \cdots,$$

cuyo intervalo de convergencia es $-1 < x < 1$.

e. Ejemplo de multiplicación de series

El desarrollo de la función

$$\frac{\log(1 + x)}{1 + x}$$

es un ejemplo sencillo de la aplicación de la regla para multiplicar dos series de potencias. Sólo tiene que multiplicarse la serie logarítmica

$$\log(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + - \cdots$$

por la serie geométrica

$$\frac{1}{1 + x} = 1 - x + x^2 - x^3 + x^4 - + \cdots.$$

Como el lector puede fácilmente verificar, resulta el notable desarrollo

$$\frac{\log(1 + x)}{1 + x} = x - (1 + \tfrac{1}{2})x^2 + (1 + \tfrac{1}{2} + \tfrac{1}{3})x^3 - (1 + \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{4})x^4 + - \cdots$$

válido para $|x| < 1$.

f. Ejemplo de integración término a término (integral elíptica)

En aplicaciones anteriores (pp. 320, 430) hemos encontrado la integral elíptica

$$K = \int_0^{\pi/2} \frac{d\phi}{\sqrt{(1 - k^2 \,\text{sen}^2 \phi)}}; \quad \text{para } (k^2 < 1)$$

que da el período de oscilación de un péndulo. Para evaluar la integral, puede primero desarrollarse el integrando mediante el teorema de la serie del binomio. Tal desarrollo da

$$\frac{1}{\sqrt{(1-k^2\operatorname{sen}^2\phi)}} = 1 + \tfrac{1}{2}k^2\operatorname{sen}^2\phi + \frac{1\cdot 3}{2\cdot 4}k^4\operatorname{sen}^4\phi + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}k^6\operatorname{sen}^6\phi + \cdots.$$

Puesto que $k^2\operatorname{sen}^2\phi$ nunca es mayor que k^2, esta serie converge uniformemente para todos los valores de ϕ, y de esta manera es posible integrar término a término:

$$K = \int_0^{\pi/2}\frac{d\phi}{\sqrt{(1-k^2\operatorname{sen}^2\phi)}} = \int_0^{\pi/2} d\phi + \frac{1}{2}k^2\int_0^{\pi/2}\operatorname{sen}^2\phi\,d\phi + \frac{1\cdot 3}{2\cdot 4}k^4\int_0^{\pi/2}\operatorname{sen}^4\phi\,d\phi + \cdots.$$

Las integrales que aquí aparecen ya han sido calculadas [ver Ec. (76), p. 300]. Substituyendo sus valores se obtiene

$$K = \int_0^{\pi/2}\frac{d\phi}{\sqrt{(1-k^2\operatorname{sen}^2\phi)}} = \frac{\pi}{2}\left[1 + \left(\frac{1}{2}\right)^2 k^2 + \left(\frac{1\cdot 3}{2\cdot 4}\right)^2 k^4 + \left(\frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\right)^2 k^6 + \cdots\right].$$

7.7 Series de potencias con términos complejos

a. Introducción de términos complejos en una serie de potencias. Representaciones complejas de funciones trigonométricas

La semejanza entre ciertas series de potencias representativas de funciones sin relación aparente entre sí condujo a Euler a establecer una conexión puramente formal entre dichas series, asignando valores complejos y, en particular, valores imaginarios puros, a la variable x. Primero describiremos el descubrimiento de Euler, formal pero por demás fructífero y notable, dejando de lado las cuestiones del rigor lógico. Después indicaremos una justificación más rigurosa.

La primer relación de esta clase se obtiene al reemplazar la cantidad x en la serie para e^x por una cantidad imaginaria pura, $i\phi$, donde ϕ es un número real. Teniendo en cuenta la ecuación fundamental para la unidad imaginaria i, esto es, $i^2 = -1$, de la cual se obtienen $i^3 = -1$, $i^4 = -1$, $i^5 = i, \ldots$, se ve que al separar los términos reales e imaginarios de la serie resulta

$$e^{i\phi} = \left(1 - \frac{\phi^2}{2!} + \frac{\phi^4}{4!} - \frac{\phi^6}{6!} + - \cdots\right) + i\left(\phi - \frac{\phi^3}{3!} + \frac{\phi^5}{5!} - \frac{\phi^7}{7!} + - \cdots\right),$$

o, en otra forma

$$(11) \qquad e^{i\phi} = \cos\phi + i \operatorname{sen}\phi.$$

Esta es la bien conocida e importante "fórmula de Euler", fundamental en el Análisis, la que, por ahora, no es sino una relación puramente formal.[1] Está de acuerdo con el teorema de De Moivre (p. 128), expresado por la ecuación

$$(\cos\phi + i \operatorname{sen}\phi)(\cos\psi + i \operatorname{sen}\psi) = \cos(\phi + \psi) + i \operatorname{sen}(\phi + \psi).$$

A la luz de la fórmula de Euler, esta ecuación simplemente establece que la relación

$$e^x \cdot e^y = e^{x+y}$$

sigue siendo válida para valores imaginarios puros, $x = i\phi$, $y = i\psi$.

Debe dejarse establecido que esta fórmula de Euler y el teorema de adición $e^{i\varphi}e^{i\psi} = e^{i(\varphi+\psi)}$ pueden ser usados rigurosamente, sin necesidad de justificación posterior, *definiendo* simplemente $e^{i\phi}$ como el número complejo $\cos\phi + i \operatorname{sen}\phi$. Esta definición es compatible con las reglas ordinarias para operar con exponenciales. En particular, la regla ordinaria para multiplicar potencias de e proporciona justamente expresiones concisas y simples de los teoremas de adición de la trigonometría expresados por la fórmula de De Moivre, la que es de un carácter enteramente elemental. Por tanto, podremos usar las relaciones de Euler con la confianza de estar pisando terreno firme, sin necesidad de recurrir a un análisis más general de las funciones de una variable compleja, cosa que haremos en la siguiente sección.

En forma más general, la función exponencial puede ser definida para un exponente complejo arbitrario $x + iy$ (donde x e y son reales) mediante la fórmula

$$e^{x+iy} = e^x e^{iy} = e^x(\cos y + i \operatorname{sen} y).$$

Al reemplazar en la serie de potencias para $\cos x$ la variable x por la cantidad imaginaria pura ix, se obtiene de inmediato la serie para $\cosh x$. Esta relación queda expresada por la ecuación

$$(12) \qquad \cosh x = \cos ix.$$

De manera similar se obtiene

$$(13) \qquad \operatorname{senh} x = \frac{1}{i} \operatorname{sen} ix.$$

Puesto que la fórmula de Euler también da $e^{-i\phi} = \cos\phi - i \operatorname{sen}\phi$, llegamos a las *expresiones exponenciales para las funciones trigonométricas:*

$$(14) \qquad \operatorname{sen} x = \frac{e^{ix} - e^{-ix}}{2i}, \qquad \cos x = \frac{e^{ix} + e^{-ix}}{2}.$$

[1] Una consecuencia que resulta para $\phi = \pi$ es la fórmula $e^{\pi i} = -1$, notable relación entre tres de los números más importantes, e, π e i.

Estas expresiones son análogas a las expresiones exponenciales de las funciones hiperbólicas y, en efecto, se transforman en éstas mediante las relaciones $x = \cos ix$, $\operatorname{sen} x = \frac{1}{i} \operatorname{sen} ix$.

Pueden también obtenerse, por supuesto, las correspondientes relaciones formales para las funciones $\tan x$, $\tanh x$, $\cot x$, $\coth x$, relacionadas entre sí por las ecuaciones $x = \frac{1}{i} \tan ix$, $\coth x = i \cot ix$.

Finalmente, diremos que es posible encontrar también relaciones análogas para las funciones trigonométricas inversas y las hiperbólicas. Por ejemplo, a partir de

$$y = \tan x = \frac{e^{ix} - e^{-ix}}{i(e^{ix} + e^{-ix})} = \frac{e^{2ix} - 1}{i(e^{2ix} + 1)}$$

se encuentra inmediatamente que

$$e^{2ix} = \frac{1 + iy}{1 - iy}.$$

Tomando logaritmos en ambos lados de esta ecuación y luego escribiendo x en vez de y y $\operatorname{arc} \tan x$ en vez de x, se obtiene la ecuación

$$(15) \qquad \operatorname{arc} \tan x = \frac{1}{2i} \log \frac{1 + ix}{1 - ix},$$

la cual expresa una relación notable entre la tangente inversa y el logaritmo. Si en la serie de potencias para $\frac{1}{2} \log \frac{1 + x}{1 - x}$ (p. 444), ya conocida, se reemplaza x por ix, se obtiene la serie de potencias para $\operatorname{arc} \tan x$,

$$\begin{aligned} \operatorname{arc} \tan x &= \frac{1}{i}\left(ix + \frac{(ix)^3}{3} + \frac{(ix)^5}{5} + \cdots \right) \\ &= x - \frac{x^3}{3} + \frac{x^5}{5} - + \cdots. \end{aligned}$$

Estas relaciones son, hasta este punto, de carácter puramente formal, y, naturalmente, es necesario establecer de manera más precisa el contenido de las mismas. Sin embargo, como se ha visto antes, mediante el uso de definiciones adecuadas es posible dar a esas relaciones significado satisfactoriamente riguroso.

*b. Un vistazo a la teoría general de funciones de una variable compleja

Aunque el punto de vista puramente formal indicado en la última sección está libre de objeciones, es deseable, no obstante, reconocer en las fórmulas anteriores algo de carácter más fundamental que simples relaciones formales. Esta idea conduce a la *teoría general de funciones complejas,* como se denomina (por brevedad) a la teoría general de las llamadas *funciones analíticas de una variable compleja.* Como punto de partida podemos usar una discusión general de la teoría de series de potencias con variables complejas y coeficientes también complejos. La construcción de una teoría tal de series de potencias no presenta dificultades una vez que se han definido los conceptos de límite en el dominio de los números complejos. De hecho, esta teoría se desarrolla en forma casi paralela a la teoría de series de potencias reales. Sin embargo, puesto que no haremos uso alguno de estas cuestiones en lo que sigue, nos concretaremos aquí a establecer ciertos hechos, omitiendo las demostraciones. Uno de los resultados que se encuentran es la siguiente generalización del teorema de la sección 7.5*a,* válida para series de potencias complejas:

Si una serie de potencias converge para cualquier valor complejo $x = \xi$, *entonces converge absolutamente para todo valor x tal que* $|x| < |\xi|$; *y si dicha serie diverge para un valor* $x = \xi$, *entonces diverge también para todo otro valor x tal que* $|x| > |\xi|$. *Una serie de potencias que no converge en todas partes pero que converge para algunos puntos, además de hacerlo para* $x = 0$, *posee un círculo de convergencia, es decir, existe un número* $\rho > 0$ *tal que la serie converge absolutamente para* $|x| < \rho$ *y diverge para* $|x| > \rho$.

Una vez que se ha establecido el concepto de "funciones de una variable compleja representadas por series de potencias", y habiendo desarrollado las reglas para operar con tales funciones, podemos considerar las funciones e^x, $\operatorname{sen} x$, $\cos x$, $\arctan x$, etc., de la variable *compleja x* simplemente como *definidas* por las series de potencias que las representan para valores reales de x.

Ilustraremos mediante dos ejemplos la forma en que la introducción de variables complejas arroja luz sobre el comportamiento de las funciones elementales. La serie geométrica para $1/(1 + x^2)$ deja de ser convergente cuando x cae fuera del intervalo $-1 \leq x \leq 1$, y lo mismo ocurre para la serie que representa a $\arctan x$, aunque el comportamiento de estas funciones no exhibe peculiaridades en los extremos del intervalo de convergencia. Estas funciones y sus derivadas son continuas para todos los valores reales de x. Por otra parte, es fácil comprender por qué las series para

$1/(1-x^2)$ y $\log(1-x)$ dejan de ser convergentes cuando x pasa por el valor uno, ya que en este punto las funciones se vuelven infinitas.

Sin embargo, la divergencia de la serie que representa a la tangente inversa y la de la serie $\sum_{\nu=0}^{\infty}(-1)^{\nu}x^{2\nu}$, para $|x|>1$, se aclara de inmediato si consideramos valores complejos de x, como antes se hizo. En efecto, en este caso encontramos que cuando $x=i$ las funciones se vuelven infinitas y por esta razón no pueden ser representadas mediante series convergentes. Por tanto, como una consecuencia del teorema sobre el círculo de convergencia, la serie debe ser divergente para todos los valores de x tales que $|x|>|i|=1$. En particular, para valores reales de x la serie diverge fuera del intervalo $-1\leq x\leq 1$.

Otro ejemplo lo constituye la función $f(x)=e^{-1/x^2}$ para $x\neq 0$, con $f(0)=0$ (ver p. 481), la cual, a pesar de su comportamiento completamente suave,[1] no puede ser desarrollada en serie de Taylor. De hecho, esta función deja de ser continua si consideramos valores imaginarios puros $x=i\xi$. La función toma entonces la forma e^{1/ξ^2} y crece más allá de toda cota cuando $\xi\to 0$. Es claro, por consiguiente, que esta función no puede ser representada por ninguna serie de potencias en x para todos los valores complejos de x en una vecindad del origen, sin importar cuán pequeña se escoja dicha vecindad.

Estas observaciones sobre la teoría de funciones y series de potencias de una variable compleja bastarán por ahora.

Apéndice

*A.1 Multiplicación y división de series

a. Multiplicación de series absolutamente convergentes

Sean

$$A=\sum_{\nu=0}^{\infty}a_\nu \quad \text{y} \quad B=\sum_{\nu=0}^{\infty}b_\nu$$

dos series absolutamente convergentes. Junto con éstas consideremos las correspondientes series convergentes de valores absolutos

$$\bar{A}=\sum_{\nu=0}^{\infty}|a_\nu| \quad \text{y} \quad \bar{B}=\sum_{\nu=0}^{\infty}|b_\nu|.$$

A continuación hagamos

$$A_n=\sum_{\nu=0}^{n-1}a_\nu, \quad B_n=\sum_{\nu=0}^{n-1}b_\nu, \quad \bar{A}_n=\sum_{\nu=0}^{n-1}|a_\nu|, \quad \bar{B}_n=\sum_{\nu=0}^{n-1}|b_\nu|$$

y

[1] Es una función "lisa", como se vio, en todo el campo real. *(N. del T.)*

$$c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0$$

Afirmamos que la serie $\sum_{\nu=0}^{\infty} c_n$ es absolutamente convergente y que su suma es igual a AB.

Para demostrarlo escribimos la serie

$$a_0b_0 + a_1b_0 + a_1b_1 + a_0b_1 + a_2b_0 + a_2b_1$$
$$+ a_2b_2 + a_1b_2 + a_0b_2 + \cdots + a_nb_0 + a_nb_1$$
$$+ \cdots + a_nb_n + \cdots + a_1b_n + a_0b_n + \cdots,$$

en la cual la n^2-ésima suma parcial es A_nB_n. Lo que se afirma es que tal serie converge absolutamente. En efecto, las sumas parciales de las correspondientes series de valores absolutos crecen monótonamente; y la n^2-ésima suma parcial es igual a $\bar{A}_n\bar{B}_n$, que es menor que $\bar{A}\bar{B}$ (y que tiende a $\bar{A}\bar{B}$). Por consiguiente, la serie de valores absolutos converge y la serie en cuestión converge absolutamente. La suma de la serie es evidentemente AB, puesto que la n^2-ésima suma parcial es A_nB_n, que tiende a AB cuando $n \to \infty$. Intercambiemos ahora los términos, lo cual es válido para series absolutamente convergentes, y agrupemos en paréntesis, en forma sucesiva, los términos. En una serie convergente es posible agrupar los términos sucesivos en la forma que se quiera sin alterar la convergencia o el valor de la suma de la serie, pues si, por ejemplo, se agrupan todos los términos $(a_{n+1} + a_{n+2} + \cdots + a_m)$, entonces al formar las sumas parciales omitiremos aquéllas sumas que originalmente caían entre s_n y s_m, lo cual no afecta la convergencia ni cambia el valor del límite. También, si antes de insertar los paréntesis la serie es absolutamente convergente, lo sigue siendo después. Puesto que la serie

$$\sum_{\nu=0}^{\infty} c_\nu = (a_0b_0) + (a_0b_1 + a_1b_0) + (a_0b_2 + a_1b_1 + a_2b_0\ ) + \cdots$$

está formada de esta manera por las series consideradas antes, la demostración requerida queda completa.

*b. Multiplicación y división de series de potencias

La aplicación principal del teorema anterior tiene lugar en la teoría de las series de potencias. El siguiente aserto es una consecuencia inmediata de él: El producto de dos series de potencias,

$$\sum_{\nu=0}^{\infty} a_\nu x^\nu \quad \text{y} \quad \sum_{\nu=0}^{\infty} b_\nu x^\nu;$$

está representado en el intervalo de convergencia común a ambas series de potencias por una tercera serie, $\sum_{\nu=0}^{\infty} c_\nu x^\nu$, cuyos coeficientes están dados por

$$c_\nu = a_0b_\nu + a_1b_{\nu-1} + \cdots + a_\nu b_0.$$

*Como en el caso de la multiplicación de series de potencias, el cociente de las dos series anteriores puede representarse por una serie de potencias $\sum_{\nu=0}^{\infty} q_\nu x^\nu$, siempre que b_0, el término constante en el denominador, no se anule. (En este último caso tal representación es por lo general imposible, pues podría no haber convergencia para $x = 0$ debido a la anulación del denominador, en tanto que, por otra parte, toda serie de potencias debe converger para $x = 0$.) Los coeficientes de la serie de potencias

$$\sum_{\nu=0}^{\infty} q_\nu x^\nu$$

pueden calcularse recordando que $\sum_{\nu=0}^{\infty} q_\nu x^\nu \cdot \sum_{\nu=0}^{\infty} b_\nu x^\nu = \sum_{\nu=0}^{\infty} a_\nu x^\nu$, y que, por consiguiente, son válidas las siguientes ecuaciones:

$$
\begin{aligned}
a_0 &= q_0 b_0, \\
a_1 &= q_0 b_1 + q_1 b_0, \\
a_2 &= q_0 b_2 + q_1 b_1 + q_2 b_0, \\
&\cdots\cdots\cdots\cdots\cdots \\
a_\nu &= q_0 b_\nu + q_1 b_{\nu-1} + \cdots + q_\nu b_0.
\end{aligned}
$$

De la primera de estas ecuaciones fácilmente se encuentra q_0, de la segunda se obtiene el valor de q_1, de la tercera (usando los valores de q_0 y q_1) se encuentra el valor de q_2, etc. Para dar una justificación estricta de la expresión del cociente de dos series de potencias como una tercera serie, hemos de investigar la convergencia de la serie de potencias $\sum_{\nu=0}^{\infty} q_\nu x^\nu$, que se ha calculado formalmente. Sin embargo, no haremos uso posterior del resultado y nos conformaremos con mencionar que la serie para el cociente realmente converge en algún intervalo en torno al origen. Se omite la demostración.

A.2 Series infinitas e integrales impropias

Las series infinitas y los conceptos desarrollados en relación con ellas tienen aplicaciones y analogías simples en la teoría de las integrales impropias (ver capítulo 3, p. 320). Nos concretamos aquí al caso de una integral convergente con un intervalo de integración infinito; digamos, una integral de la forma $\int_0^\infty f(x)\,dx$. Si dividimos el intervalo de integración mediante la sucesión de números $x_0 = 0$, x_1, ... que tiende monótonamente a $+\infty$, podemos escribir la integral impropia en la forma

$$\int_0^\infty f(x)\,dx = a_1 + a_2 + \cdots,$$

donde cada término de la serie infinita es una integral:

$$a_1 = \int_0^{x_1} f(x)\,dx, \qquad a_2 = \int_{x_2}^{x_1} f(x)\,dx, \ldots,$$

y así sucesivamente. Esto es válido independientemente de la forma en que se escojan los puntos x_ν. Podemos, por consiguiente, relacionar la idea de una integral impropia convergente con la de una serie infinita, de muchas maneras.

Es especialmente conveniente escoger los puntos x_ν de manera tal que el integrando no cambie de signo dentro de ninguno de los subintervalos individuales. La serie $\sum_{\nu=1}^{\infty} |a_\nu|$ corresponde entonces a la integral del valor absoluto de nuestra función,

$$\int_0^\infty |f(x)|\, dx.$$

Hemos así llegado, de manera natural, al siguiente concepto: *se dice que una integral impropia,* $\int_0^\infty f(x)\, dx$, *es absolutamente convergente si la integral* $\int_0^\infty |f(x)|\, dx$ *converge.* De lo contrario, si la primera integral existe y la segunda no, decimos que aquélla es *condicionalmente convergente.*

Algunas de las integrales antes consideradas (pp. 326 a 328), tales como

$$\int_0^\infty \frac{1}{1+x^2}\, dx, \qquad \int_0^\infty e^{-x^2}\, dx, \qquad \Gamma(x) = \int_0^\infty e^{-t}t^{x-1}\, dt,$$

son absolutamente convergentes.

Por otra parte, la importante integral de Dirichlet

$$J = \int_0^\infty \frac{\operatorname{sen} x}{1}\, dx = \lim_{A\to\infty} \int_0^A \frac{\operatorname{sen} x}{x}\, dx,$$

estudiada en la p. 328, es el ejemplo típico de una integral condicionalmente convergente. La demostración más sencilla posible de su convergencia se basa en la reducción a una integral absolutamente convergente: Escribimos $\operatorname{sen} x = (1 - \cos x)' = 2(\operatorname{sen}^2 x/2)'$ y aplicamos integración por partes, con lo cual J se reduce a la forma absolutamente convergente

$$J = 2\int_\theta^\infty \left(\operatorname{sen}^2 \frac{x}{2}\right)\frac{1}{x^2}\, dx$$

(Obsérvese que el nuevo integrando tiende continuamente al límite $\frac{1}{2}$ cuando $x \to 0$ y se anula con el orden de x^{-2} cuando $x \to \infty$,)

*Una demostración diferente de la convergencia se obtiene subdividiendo el intervalo de 0 a A mediante los puntos $x_\nu = \nu\pi (\nu = 0, 1, 2, \ldots, \mu_A)$, donde μ_A es el mayor entero posible que satisface $\mu_A\pi \leq A$. Por consiguiente, la integral resulta dividida en términos de la forma

$$a_\nu = \int_{(\nu-1)\pi}^{\nu\pi} \frac{\operatorname{sen} x}{x}\, dx, \qquad \text{para} \qquad \nu = 1, 2, \ldots,$$

y un residuo R_A de la forma

$$\int_{\mu_A\pi}^{A} \frac{\operatorname{sen} x}{x}\, dx \qquad (0 \leq A - \mu_A\pi < \pi).$$

Es obvio que las cantidades a_ν tienen signos alternantes, pues $\operatorname{sen} x$ es alternadamente positivo y negativo en intervalos consecutivos. Además, $|a_{\nu+1}| < |a_\nu|$, pues al aplicar la transformación $x = \xi - \pi$, se tiene

$$|a_\nu| = \int_{(\nu-1)\pi}^{\nu\pi} \frac{|\text{sen}\, x|}{x}\, dx = \int_{\nu\pi}^{(\nu+1)\pi} \frac{|\text{sen}\,(\xi - \pi)|}{\xi - \pi}\, d\xi = \int_{\nu\pi}^{(\nu+1)\pi} \frac{|\text{sen}\,\xi|}{\xi - \pi}\, d\xi >$$

$$\int_{\nu\pi}^{(\nu+1)\pi} \frac{|\text{sen}\,\xi|}{\xi}\, d\xi = |a_{\nu+1}|.$$

En consecuencia, por el criterio de Leibnitz se ve que Σa_ν converge. Además, el residuo R_A tiene el valor absoluto

$$|R_A| = \left| \int_{\mu_A\pi}^{A} \frac{\text{sen}\, x}{x}\, dx \right| \leq \int_{\mu_A\pi}^{(\mu_A+1)\pi} \frac{|\text{sen}\, x|}{x}\, dx$$

$$\leq \frac{1}{\mu_A\pi} \int_{\mu_A\pi}^{(\mu_A+1)\pi} |\text{sen}\, x|\, dx = \frac{2}{\mu_A\pi},$$

que tiende a 0 cuando A crece. Así, si se hace que A tienda a infinito en la ecuación

$$\int_{0}^{A} \frac{\text{sen}\, x}{x}\, dx = a_1 + a_2 + a_3 + \cdots + a_{\mu_A} + R_A,$$

el lado derecho tiende a Σa_ν como límite, y así la integral es convergente. Sin embargo, la convergencia no es absoluta para

$$|a_\nu| > \int_{(\nu-1)\pi}^{\nu\pi} \frac{|\text{sen}\, x|}{\nu\pi}\, dx = \frac{2}{\nu\pi}, \qquad \text{de manera que } \sum |a_\nu| \text{ diverge.}$$

*A.3 Productos infinitos

En la introducción a este capítulo (p. 530) se dijo que las series infinitas son sólo *una* forma, aunque particularmente importante, de representar números o funciones mediante procesos infinitos. Como un ejemplo de otro de tales procesos consideraremos aquí los productos infinitos. Pero no daremos demostraciones.

En la p. 301 encontramos el producto de Wallis,

$$\frac{\pi}{2} = \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdots,$$

en el cual el número $\pi/2$ se expresa como un "producto infinito". Generalmente hablando, por *valor* del producto infinito

$$\prod_{\nu=1}^{\infty} a_\nu = a_1 \cdot a_2 \cdot a_3 \cdot a_4 \cdots$$

se entiende el *límite* de la sucesión de "productos parciales"

$$a_1, \quad a_1 \cdot a_2, \quad a_1 \cdot a_2 \cdot a_3, \quad a_1 \cdot a_2 \cdot a_3 \cdot a_4, \ldots,$$

en el supuesto de que exista.

Es claro que los factores $a_1, a_2, a_3, \ldots$ pueden ser también funciones de una variable x. Un ejemplo especialmente interesante es el del "producto infinito" para la función sen x,

$$\text{sen}\,\pi x = \pi x\left(1-\frac{x^2}{1^2}\right)\left(1-\frac{x^2}{2^2}\right)\left(1-\frac{x^2}{3^2}\right)\cdots, \tag{16}$$

que obtendremos en la sección 8.5, p. 620.

El producto infinito para la *función zeta* juega un papel muy importante en la teoría de los números. Para conservar la notación usual de la teoría de los números, denotaremos aquí por s la variable independiente y definiremos la función zeta para $s > 1$, siguiendo a Riemann, por la expresión

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

Sabemos (sección 7.2*c*, p. 543) que la serie de la derecha converge cuando $s > 1$. Si p es cualquier número mayor que 1, se obtiene la ecuación

$$\frac{1}{1-\dfrac{1}{p^s}} = 1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \frac{1}{p^{3s}} + \cdots$$

al desarrollar el lado de la izquierda en una serie geométrica de cociente p^{-s}. Imaginemos que esta serie se escribe para todos los números primos $p_1, p_2, p_3, \ldots$ considerados en orden de magnitud creciente, y que todas las ecuaciones que así resultan se multiplican entre sí. Obtenemos en el lado izquierdo un producto de la forma

$$\frac{1}{1-p_1^{-s}}\cdot\frac{1}{1-p_2^{-s}}\cdots.$$

Sin detenernos a justificar este procedimiento multipliquemos entre sí las series del lado derecho de nuestras ecuaciones. Como resultado se obtiene una suma de términos

$$p_1^{-k_1 s}p_2^{-k_2 s}p_3^{-k_3 s}\cdots = (p_1^{k_1}p_2^{k_2}p_3^{k_3}\cdots)^{-s},$$

donde $k_1, k_2, k_3, \ldots$ son enteros no negativos. Recordamos también que, por un teorema elemental, todo entero $n > 1$ puede ser expresado de manera única como un producto de potencias de números primos diferentes: $n = p_1^{k_1}p_2^{k_2}\cdots$. En esta forma encontramos que el producto del lado derecho es la función $\zeta(s)$, de manera que obtenemos la notable "fórmula del producto" de Euler:

$$\zeta(s) = \frac{1}{1-p_1^{-s}}\cdot\frac{1}{1-p_2^{-s}}\cdot\frac{1}{1-p_3^{-s}}\cdots. \tag{17}$$

Esta "fórmula del producto", cuya derivación sólo se ha bosquejado aquí, es en realidad una expresión de la función zeta como un producto infinito, puesto que el número de números primos es infinito.

En la teoría general de los productos infinitos se excluye, por lo general, el caso en que el producto $a_1a_2\cdots a_n$ tiene como límite el cero. Por consiguiente, es especialmente importante que ninguno de los factores a_n se anule. Para que el producto pueda ser convergente los factores a_n deben, según esto, tender a 1 cuando n crece; y, puesto que, si es necesario, pode-

mos omitir un número finito de factores (lo cual no influye en las cuestiones de convergencia), es válido suponer que $a_n > 0$. El siguiente teorema, casi trivial, es aplicable a este caso:

Una condición necesaria y suficiente para la convergencia del producto $\prod_{\nu=1}^{\infty} a_\nu$, *donde* $a_\nu > 0$, *es que la serie* $\sum_{\nu=1}^{\infty} \log a_\nu$ *sea convergente.* En efecto, como una consecuencia de la continuidad del logaritmo, las sumas parciales $\sum_{\nu=1}^{n} \log a_\nu = \log (a_1 a_2 \cdots a_n)$ de esta serie tenderán a un límite definido si, y sólo si, los productos parciales $a_1 a_2 \cdots a_n$ tienen un límite positivo.

En el estudio de la convergencia es de gran utilidad el siguiente *criterio que da una condición suficiente,* en la que $a_\nu = 1 + \alpha_\nu$. *El producto*

$$\prod_{\nu=1}^{\infty} (1 + \alpha_\nu)$$

converge si la serie

$$\sum_{\nu=1}^{\infty} |\alpha_\nu|$$

converge y ninguno de los factores $(1 + \alpha_\nu)$ *es nulo.* Para la demostración podemos suponer, omitiendo un número finito de factores si es necesario, que $|\alpha_\nu| < \frac{1}{2}$ para todo ν. Entonces tenemos $1 - |\alpha_\nu| > \frac{1}{2}$ y, por el teorema del valor medio, $\log (1 + h) = \log (1 + h) - \log 1 = h/(1 + \theta h)$ con $0 < \theta < 1$.

Por consiguiente

$$|\log (1 + \alpha_\nu)| = \left| \frac{\alpha_\nu}{1 + \theta \alpha_\nu} \right| \leq \frac{|\alpha_\nu|}{1 - |\alpha_\nu|} \leq 2\, |\alpha_\nu|,$$

y en esta forma la convergencia de la serie $\sum_{\nu=1}^{\infty} \log (1 + \alpha_\nu)$ se deduce de la convergencia de $\sum_{\nu=1}^{\infty} |\alpha_\nu|$.

De este criterio se deduce que el producto infinito considerado antes para sen πx, (16), converge para todos los valores de x excepto para los valores $x = 0, \pm 1, \pm 2, \ldots$, en los cuales los factores del producto se anulan. Lo mismo que para la función ζ de Riemann, cuando $p \geq 2$ y $s > 1$ se encuentra de inmediato que

$$\frac{1}{1 - p^{-s}} = 1 + \frac{1}{p^s - 1}, \qquad 0 < \frac{1}{p^s - 1} < \frac{2}{p^s}.$$

Ahora bien, si p toma todos los valores primos, la serie $\Sigma \frac{1}{p^s}$ debe converger, puesto que sus términos son sólo una parte de la serie convergente $\sum_{\nu=1}^{\infty} \frac{1}{\nu^s}$. Queda así demostrada la convergencia para $s > 1$ del producto de la Ec. (17). Puesto que la serie

para $\zeta(s)$ con $s = 1$ (esto es, la serie armónica) diverge, puede concluirse que la serie de inversos de números primos, es decir, la serie

$$\sum_{k=1}^{\infty} \frac{1}{p_k} = \frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{11} + \frac{1}{13} + \frac{1}{17} + \frac{1}{19} + \cdots$$

diverge también. (De paso esto demuestra que el número de números primos es infinito.) En efecto, si la serie de números primos recíprocos fuese convergente, entonces también lo sería la serie con términos

$$\alpha_k = \frac{1}{1 - p_k^{-1}} - 1 = \frac{p_k^{-1}}{1 - p_k^{-1}}$$

puesto que $p_k \geq 2$ y

$$0 < \alpha_k \leq 2p_k^{-1}.$$

En consecuencia, como indica el criterio anterior, el producto infinito

$$\prod_{k=1}^{\infty} (1 + \alpha_k) = \prod_{k=1}^{\infty} \frac{1}{1 - p_k^{-1}} = \prod_{k=1}^{\infty} \left(1 + \frac{1}{p_k} + \frac{1}{p_k^2} + \cdots\right)$$

también sería convergente; pero entonces la serie armónica debería converger, lo cual es imposible.

*A.4 Series en que aparecen números de Bernoulli

Hasta ahora no hemos dado los desarrollos en series de potencias para ciertas funciones elementales; por ejemplo, para tan x. La razón de ello es que los coeficientes numéricos que aparecen en esos casos no tienen ninguna forma sencilla. Estos coeficientes, y los que aparecen en las series de ciertas funciones, pueden expresarse en términos de los llamados *números de Bernoulli.* Son éstos curiosos números racionales, con una ley de formación algo obscura, que aparecen en muchas partes del Análisis. La manera más simple de obtenerlos consiste en desarrollar la función

$$\frac{x}{e^x - 1} = \frac{1}{1 + \dfrac{x}{2!} + \dfrac{x^2}{3!} + \cdots}$$

en una serie de potencias formal del tipo

$$\frac{x}{e^x - 1} = \sum_{\nu=0}^{\infty} \frac{B_\nu^*}{\nu!} x^\nu.$$

Escribiendo esta ecuación en la forma

$$x = (e^x - 1) \sum_{\nu=0}^{\infty} \frac{B_\nu^*}{\nu!} x^\nu$$

y substituyendo en el lado derecho $e^x - 1$ por su serie de potencias, se obtiene una relación de recurrencia para los B_n^*:

$$\binom{n+1}{1} B_n^* + \binom{n+1}{2} B_{n-1}^* + \binom{n+1}{3} B_{n-2}^* + \cdots + \binom{n+1}{n+1} B_0^* = 0,$$

donde $n > 0$ y $B_0{}^* = 1$. A partir de ella es posible calcular con facilidad los $B_i{}^*$ en forma sucesiva. Estos números racionales se llaman números de Bernoulli.[1] Son números racionales ya que en su formación sólo intervienen operaciones racionales. Como es fácil ver, se anulan para todos los índices impares diferentes de $\nu = 1$. Algunos de ellos se dan a continuación

$$B_0{}^* = 1, \quad B_1{}^* = -\frac{1}{2}, \quad B_2{}^* = \frac{1}{6}, \quad B_4{}^* = -\frac{1}{30}, \quad B_6{}^* = \frac{1}{42},$$

$$B_8{}^* = -\frac{1}{30}, \quad B_{10}{}^* = \frac{5}{66}, \ldots$$

Nos conformaremos aquí con dar una idea breve de cómo estos números aparecen en las series de potencias que estamos considerando. Primero, haciendo uso de la transformación

$$1 + \frac{B_2{}^*}{2!}x^2 + \cdots = \frac{x}{e^x - 1} + \frac{x}{2} = \frac{x}{2}\cdot\frac{e^x + 1}{e^x - 1} = \frac{x}{2}\cdot\frac{e^{\frac{1}{2}x} + e^{-\frac{1}{2}x}}{e^{\frac{1}{2}x} - e^{-\frac{1}{2}x}},$$

obtenemos

$$\frac{x}{2}\coth\frac{x}{2} = \sum_{\nu=0}^{\infty} \frac{B_{2\nu}{}^*}{(2\nu)!}x^{2\nu}.$$

(Esta fórmula prueba que $B^*_{2\nu+1} = 0$ para $\nu > 0$, ya que $(x/2)\coth(x/2)$ es una función par de x.)

Si reemplazamos x por $2x$ obtenemos la serie

$$x \coth x = \sum_{\nu=0}^{\infty} \frac{2^{2\nu}B_{2\nu}{}^*}{(2\nu)!}x^{2\nu},$$

válida, como puede demostrarse, para $|x| < \pi$, y de la cual se obtiene

$$x \cot x = \sum_{\nu=0}^{\infty} (-1)^{\nu} \frac{2^{2\nu}B_{2\nu}{}^*}{(2\nu)!}x^{2\nu}, \qquad |x| < \pi,$$

al reemplazar x por $-ix$ (ver p. 567).

Por medio de la ecuación $2\cot 2x = \cot x - \tan x$ obtenemos ahora la serie

$$\tan x = \sum_{\nu=1}^{\infty} (-1)^{\nu-1} \frac{2^{2\nu}(2^{2\nu} - 1)}{(2\nu)!} B_{2\nu}{}^* x^{2\nu-1},$$

que se cumple para $|x| < \frac{\pi}{2}$.

Para una información adicional remitimos al lector al capítulo 8 y a textos más detallados.[2]

---

[1] Con una notación ligeramente diferente (pp. 638-639), la fórmula básica se escribirá

$$\frac{x}{e^x - 1} = 1 - \tfrac{1}{2}x + \sum_{\nu=1}^{\infty} (-1)^{\nu+1} \frac{B_\nu}{(2\nu)!} x^{2\nu}.$$

[2] Véase, por ejemplo, K. Knopp, *Theory and Applications of Infinite Series*, página 183, Blackie & Son, Ltd., 1928, y K Knopp, *Infinite Sequences and Series*, Dover Publications, 1956.

PROBLEMAS

SECCION 7.1, página 528

1. Demostrar que

$$\sum_{\nu=1}^{\infty} \frac{1}{\nu(\nu+1)} = \frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \cdots + = 1$$

[ver problema 1.6, 12(*a*)] y úsese el resultado para demostrar que $\sum_{\nu=1}^{\infty} \frac{1}{\nu^2}$ converge.

2. Usar el resultado del problema 1 para obtener cotas superior e inferior para

$$\sum_{\nu=1}^{\infty} \frac{1}{\nu^2}.$$

3. Demostrar que $\sum_{\nu=0}^{\infty} (-1)^{\nu} \frac{2\nu+3}{(\nu+1)(\nu+2)} = 1.$

4. ¿Para qué valores de α converge la serie $1 - \frac{1}{2^{\alpha}} + \frac{1}{3^{\alpha}} - \frac{1}{4^{\alpha}} + \cdots$?

5. Demostrar que si $\sum_{\nu=1}^{\infty} a_{\nu}$ converge y $s_n = a_1 + a_2 + \cdots + a_n$, entonces la sucesión

$$\frac{s_1 + s_2 + \cdots + s_N}{N}$$

también converge y tiene por límite $\sum_{\nu=1}^{\infty} a_{\nu}$.

6. ¿Es convergente la serie $\sum_{n=1}^{\infty} \left(\frac{2n}{2n+1} - \frac{2n-1}{2n} \right)$?

7. ¿Es convergente la serie $\sum_{\nu=1}^{\infty} (-1)^{\nu} \frac{\nu}{\nu+1}$?

8. Demostrar que si $\sum_{\nu=1}^{\infty} a_{\nu}^2$ converge, lo mismo es cierto para $\sum_{\nu=1}^{\infty} \frac{a_{\nu}}{\nu}$.

9. *a*) Si a_n es una sucesión monótona creciente con términos positivos, ¿en qué casos converge la serie $\frac{1}{a_1} + \frac{1}{a_1 a_2} + \frac{1}{a_1 a_2 a_3} + \cdots$?

b) Dar un ejemplo de una sucesión monótona decreciente tal que $\lim_{n\to\infty} a_n = 1$ y para la cual la serie asociada diverja.

c) Demostrar que si se admiten sucesiones decrecientes, es posible entonces obtener sumas convergentes aun cuando $\lim_{n\to\infty} a_n = 1$.

10. Si la serie $\sum_{\nu=1}^{\infty} a_{\nu}$ con términos positivos decrecientes converge, entonces $\lim_{n\to\infty} n a_n = 0$.

11. Demostrar que la serie $\sum_{\nu=1}^{\infty} \operatorname{sen} \frac{\pi}{\nu}$ diverge.

12. Demostrar que si Σa_ν converge y $b_1, b_2, b_3, \ldots$ es una sucesión monótona ıcotada de números, entonces $\Sigma a_\nu b_\nu$ converge. Además, demostrar que si $S = \Sigma a_\nu b_\nu$ ' $\Sigma a_\nu \leq M$, entonces $|S| \leq M b_1$.

13. Se dice que una sucesión $\{a_n\}$ es de *variación acotada* si la serie

$$\sum_{i=1}^{\infty} |a_{i+1} - a_i|$$

converge

a) Demostrar que si la sucesión $\{a_n\}$ es de variación acotada, entonces la sucesión $\{a_n\}$ converge.

b) Encontrar una serie infinita divergente, Σa_i cuyos elementos a_i constituyan una sucesión de variación acotada.

c) Demostrar la siguiente generalización del criterio de convergencia de Abel (ver página 532), debido a Dedekind:

La serie $\Sigma a_i p_i$ es convergente si Σa_i oscila entre cotas finitas y $\{p_i\}$ es una sucesión nula de variación acotada.

d) Establecer la convergencia de las siguientes series infinitas:

a) $\sum_{n=2}^{\infty} \frac{\operatorname{sen} nx}{\log n} (-1)^n$;

b) $\sum_{n=2}^{\infty} \frac{\cos nx}{\log n} (-1)^n$,

siendo x un número real arbitrario pero fijo.

14. Discutir la convergencia o divergencia de las siguientes series:

a) $\sum \frac{(-1)^\nu}{\nu}$

b) $\sum \frac{(-1)^\nu \cos(\theta/\nu)}{\nu}$

c) $\sum \frac{\cos \nu\theta}{\nu}$

d) $\sum \frac{\operatorname{sen} \nu\theta}{\nu}$

e) $\sum \frac{(-1)^\nu \cos \nu\theta}{\nu}$

f) $\sum \frac{(-1)^\nu \operatorname{sen} \nu\theta}{\nu}$

15. Encontrar las sumas de los siguientes reordenamientos de la serie para log 2,

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \cdots:$$

a) $1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + \frac{1}{5} - \frac{1}{10} - \frac{1}{12} + - - \cdots$.

b) $1 + \frac{1}{3} + \frac{1}{5} - \frac{1}{2} - \frac{1}{4} - \frac{1}{6} + + + \cdots$.

16. Determinar si las siguientes series convergen o divergen:

a) $1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} + \frac{1}{8} - \frac{1}{9} + + - \cdots$.

b) $1 + \frac{1}{2} - \frac{2}{3} + \frac{1}{4} + \frac{1}{5} - \frac{2}{6} + \frac{1}{7} + \frac{1}{8} - \frac{2}{9} + + - \cdots$.

SECCION 7.2, página 539

1. Demostrar que $\sum_{\nu=2}^{\infty} \frac{1}{\nu(\log \nu)^{\alpha}}$ converge cuando $\alpha > 1$ y diverge cuando $\alpha \leq 1$.

2. Demostrar que $\sum_{\nu=3}^{\infty} \frac{1}{\nu \log \nu (\log \log \nu)^{\alpha}}$ converge cuando $\alpha > 1$ y diverge cuando $\alpha \leq 1$

3. Demostrar que si n es un entero arbitrario mayor que 1,

$$\sum_{\nu=1}^{\infty} \frac{a_{\nu}^{n}}{\nu} = \log n,$$

donde a_{ν}^{n} se define como sigue:

$$a_{\nu}^{n} = \begin{cases} 1 \text{ si } n \text{ no es un factor de } \nu \\ -(n-1) \text{ si } n \text{ es un factor de } \nu \end{cases}$$

4. Demostrar que $\sum_{\nu=2}^{\infty} \frac{\log(\nu+1) - \log \nu}{(\log \nu)^2}$ converge.

5. Demostrar que $\sum_{\nu=1}^{\infty} \frac{1 \cdot 2 \cdot 3 \cdots \nu}{(\alpha+1)(\alpha+2)\cdots(\alpha+\nu)}$ converge si $\alpha > 1$ y diverge si $\alpha \leq 1$.

***6.** Mediante comparación con la serie $\sum_{\nu=1}^{\infty} \frac{1}{\nu^{\alpha}}$, establecer el criterio siguiente: Si $\frac{\log(1/|a_n|)}{\log n} > 1 + \varepsilon$ para algún número fijo $\varepsilon > 0$ independiente de n, y para todo n suficientemente grande, entonces la serie Σa_{ν} converge absolutamente. Si $\frac{\log(1/|a_n|)}{\log n} < 1 - \varepsilon$ para todo n suficientemente grande y algún número $\varepsilon > 0$ independiente de n, entonces la serie Σa_{ν} no converge absolutamente.

7. Demostrar que la serie $\sum_{\nu=1}^{\infty} \left(1 - \frac{1}{\sqrt{\nu}}\right)^{\nu}$ converge.

8. ¿Para qué valores de α convergen las siguientes series

a) $1 - \frac{1}{2^{\alpha}} + \frac{1}{3} - \frac{1}{4^{\alpha}} + \frac{1}{5} - \frac{1}{6^{\alpha}} + - \cdots,$

b) $1 + \frac{1}{3^{\alpha}} - \frac{1}{2^{\alpha}} + \frac{1}{5^{\alpha}} + \frac{1}{7^{\alpha}} - \frac{1}{4^{\alpha}} + - \cdots ?$

9. Por comparación con la serie $\Sigma \frac{1}{\nu(\log \nu)^{\alpha}}$ establecer el criterio siguiente: La serie $\Sigma |a_{\nu}|$ converge o diverge según que

$$\frac{\log(1/n\,|a_n|)}{\log \log n}$$

sea mayor que $1 + \varepsilon$ o menor que $1 - \varepsilon$, para todo n suficientemente grande.

10. Obtener del criterio del problema 6 el criterio de la raíz n-ésima.

11. Establecer el siguiente criterio de comparación: Si la serie Σb_ν de términos positivos converge y si

$$\left|\frac{a_{n+1}}{a_n}\right| < \frac{b_{n+1}}{b_n}$$

a partir de cierto término, la serie Σa_ν es absolutamente convergente. Si Σb_ν diverge y si

$$\left|\frac{a_{n+1}}{a_n}\right| > \frac{b_{n+1}}{b_n}$$

a partir de cierto término, la serie Σa_ν no es absolutamente convergente.

***12.** Por comparación con la serie $\sum_{\nu=1}^{\infty} \frac{1}{\nu^\alpha}$, establecer el *criterio de Raabe:*
La serie $\Sigma\,|a_\nu|$ converge o diverge según que

$$n\left(\frac{|a_n|}{|a_{n+1}|} - 1\right)$$

sea mayor que $1 + \varepsilon$ o menor que $1 - \varepsilon$ para todo n suficientemente grande y para algún $\varepsilon > 0$ independiente de n.

13. Por comparación con $\Sigma \frac{1}{\nu(\log \nu)^\alpha}$, establecer el siguiente criterio:
La serie $\Sigma\,|a_\nu|$ converge o diverge según que

$$n \log n\left(\frac{|a_n|}{|a_{n+1}|} - 1 - \frac{1}{n}\right)$$

sea mayor que $1 + \varepsilon$ o menor que $1 - \varepsilon$ para todo n suficientemente grande.

14. Probar el *criterio de Gauss:*

Si
$$\frac{|a_n|}{|a_{n+1}|} = 1 + \frac{\mu}{n} + \frac{R_n}{n^{1+\varepsilon}},$$

donde $|R_n|$ es acotado y $\varepsilon > 0$ es independiente de n, la serie $\Sigma\,|a_\nu|$ converge si $\mu > 1$, y diverge si $\mu \leq 1$.

15. Investigar la convergencia o divergencia de las siguientes series "hipergeométricas":

$$a)\quad \frac{\alpha}{\beta} + \frac{\alpha(\alpha+1)}{\beta(\beta+1)} + \frac{\alpha(\alpha+1)(\alpha+2)}{\beta(\beta+1)(\beta+2)} + \cdots.$$

$$b)\quad 1 + \frac{\alpha\cdot\beta}{1\cdot\gamma} + \frac{\alpha(\alpha+1)\cdot\beta(\beta+1)}{1\cdot 2\cdot\gamma(\gamma+1)} + \frac{\alpha(\alpha+1)(\alpha+2)\cdot\beta(\beta+1)(\beta+2)}{1\cdot 2\cdot 3\cdot\gamma(\gamma+1)(\gamma+2)} + \cdots.$$

SECCION 7.4, página 545

1. La sucesión $f_n(x)$, $n = 1, 2, \ldots$, está definida en el intervalo $0 \leq x \leq 1$ por las ecuaciones

$$f_0(x) \equiv 1, \qquad f_n(x) = \sqrt{x f_{n-1}(x)}.$$

a) Demostrar que en el intervalo $0 \leq x \leq 1$ la sucesión converge a un límite continuo.
******b*) Probar que la convergencia es uniforme.

*2. Sea $f_0(x)$ una función continua en el intervalo $0 \leq x \leq a$, y sea la sucesión de funciones $f_n(x)$ definida por

$$f_n(x) = \int_0^x f_{n-1}(t)\, dt, \qquad n = 1, 2, \ldots .$$

Demostrar que en cualquier intervalo fijo $0 \leq x \leq a$ la sucesión converge uniformemente a 0.

*3. Sea $f_n(x)$, $n = 1, 2, \ldots$, una sucesión de funciones con derivadas continuas en el intervalo $a \leq x \leq b$. Demostrar que si $f_n(x)$ converge en cada punto del intervalo y que si la desigualdad $|f_n'(x)| < M$ (donde M es una constante) se satisface para todos los valores de n y de x, entonces la convergencia es uniforme.

4. *a*) Demostrar que la serie $\sum_{\nu=1}^{\infty} \frac{1}{\nu^x}$ converge uniformemente para $x \geq 1 + \varepsilon$, siendo $\varepsilon > 0$ cualquier número fijo.

b) Demostrar que la serie derivada $-\sum \frac{\log \nu}{\nu^x}$ converge uniformemente para $x \geq 1 + \varepsilon$, donde ε es un número positivo fijo.

*5. Demostrar que la serie $\sum \frac{\cos \nu x}{\nu^\alpha}$, $\alpha > 0$, converge uniformemente para $\varepsilon \leq x \leq 2\pi - \varepsilon$ donde ε es cualquier valor positivo suficientemente pequeño.

6. La serie

$$\frac{x-1}{x+1} + \frac{1}{3}\left(\frac{x-1}{x+1}\right)^3 + \frac{1}{5}\left(\frac{x-1}{x+1}\right)^5 + \cdots$$

converge uniformemente para $\varepsilon \leq x \leq N$, siendo ε y N números positivos fijos.

7. Encontrar en qué regiones son convergentes las siguientes series:

a) $\sum x^{\nu!}$.

b) $\sum \frac{(\nu!)^2 x^\nu}{(2\nu)!}$.

c) $\sum \frac{a^\nu}{\nu^x}$, $a < 1$.

d) $\sum \frac{a^\nu}{\nu^x}$, $a > 1$.

e) $\sum \frac{\log \nu}{\nu^x}$.

f) $\sum \frac{x^\nu}{1 - x^\nu}$.

*8. Demostrar que si la serie de Dirichlet, $\sum \frac{a_\nu}{\nu^x}$, converge para $x = x_0$, converge también para cualquier $x > x_0$; y si diverge para $x = x_0$, diverge también para cualquier $x < x_0$. Existe así una "abscisa de convergencia" tal que para cualquier valor de x mayor la serie converge y para cualquier valor de x menor la serie diverge.

9. Si $\sum \frac{a_\nu}{\nu^x}$ converge para $x = x_0$, la serie derivada, $-\sum \frac{a_\nu \log \nu}{\nu^x}$, converge para cualquier $x > x_0$.

SECCION 7.5, página 556

1. Si el intervalo de convergencia de la serie de potencias $\Sigma a_n x^n$ es $|x| < \rho$, y el de $\Sigma b_n x^n$ es $|x| < \rho'$, siendo $\rho < \rho'$, ¿cuál es el intervalo de convergencia de $\Sigma(a_n + b_n)x^n$?

2. Si $a_\nu > 0$ y Σa_ν converge, demostrar que entonces

$$\lim_{x \to 1-0} \sum a_\nu x^\nu = \sum a_\nu.$$

3. Si $a_\nu > 0$ y Σa_ν diverge, demostrar que entonces

$$\lim_{x \to 1-0} \sum a_\nu x^\nu = \infty.$$

***4.** Demostrar el teorema de Abel:
Si $\Sigma a_\nu X^\nu$ converge, entonces $\Sigma a_\nu x^\nu$ converge uniformemente para $0 \leq x \leq X$.

***5.** Si $\Sigma a_\nu X^\nu$ converge, entonces $\lim_{x \to X-0} \Sigma a_\nu x^\nu = \Sigma a_\nu X^\nu$. Demostrarlo.

***6.** Por multiplicación de series de potencias demostrar que
a) $e^x e^y = e^{x+y}$. *b*) $\operatorname{sen} 2x = 2 \operatorname{sen} x \cos x$.

7. Usando la serie binomial calcular $\sqrt{2}$ con cuatro cifras decimales.

8. Sea a_n una sucesión de números reales y S el conjunto de todos los puntos límite de a_n. Denótese el extremo superior p de S por $p = \overline{\lim}\, a_n$. Mostrar que la serie de potencias $\sum_{n=0}^{\infty} c_n x^n$ converge para $|x| < \rho$ y diverge para $|x| > \rho$, donde

$$\rho = \frac{1}{\overline{\lim} \sqrt[n]{|c_n|}}.$$

APENDICE, página 570

1. Demostrar que la serie de potencias para $\sqrt{(1 - x)}$ sigue siendo convergente para $x = 1$.

2. Demostrar que para todo número positivo ε existe un polinomio en x que representa a $\sqrt{(1 - x)}$ en el intervalo $0 \leq x \leq 1$ con un error menor que ε.

3. Haciendo $x = 1 - t^2$ en el problema 2, demostrar que para todo número positivo ε existe un polinomio en t que representa a $|t|$ en el intervalo $-1 \leq t \leq 1$ con un error menor que ε.

4. *a*) Demostrar que si $f(x)$ es continua para $a \leqq x \leqq b$, entonces para todo $\varepsilon > 0$ existe una función poligonal $\varphi(x)$ (esto es, una función continua cuya gráfica consiste en un número finito de segmentos rectilíneos unidos entre sí formando esquinas) tal que $|f(x) - \varphi(x)| < \varepsilon$ para todo valor de x en el intervalo.
b) Probar que toda función poligonal $\varphi(x)$ puede ser representada mediante una suma $\varphi(x) = a + bx + \Sigma c_i |x - x_i|$, donde las x_i son las abscisas de las esquinas.

5. Teorema de aproximación de Weierstrass. Demostrar, sobre la base del aserto anterior, que si $f(x)$ es continua en $a \leq x \leq b$, entonces para todo número positivo ε existe un polinomio $P(x)$ tal que $|f(x) - P(x)| < \varepsilon$ para todos los valores de x en el intervalo $a \leq x \leq b$.
Sugerencia: Aproxímese $f(x)$ por una combinación lineal de la forma $(x - x_r) + |x - x_r|$.

6. Demostrar que los siguientes productos infinitos convergen:

a) $\prod_{n=1}^{\infty} (1 + (\frac{1}{2})^{2n})$;

b) $\prod_{n=2}^{\infty} \frac{n^3 - 1}{n^3 + 1}$;

c) $\prod_{n=1}^{\infty}\left(1-\frac{z^2}{n!}\right)$,

si $|z| < 1$.

7. Probar, usando los métodos del texto, que $\prod_{n=1}^{\infty}\left(1+\frac{1}{n}\right)$ diverge.

8. Demostrar la identidad

$$\prod_{\nu=1}^{\infty}(1+x^{2\nu})=\frac{1}{1-x}$$

para $|x| < 1$.

***9.** Considérense todos los números naturales cuya representación en el sistema decimal no contenga el número 9 entre sus dígitos. Demostrar que la suma de los inversos de esos números converge.

10. *a*) Demostrar que para $s > 1$ es

$$1-\frac{1}{2^s}+\frac{1}{3^s}-\frac{1}{4^s}+\cdots=(1-2^{1-s})\zeta(s),$$

donde $\zeta(s)$ es la función Zeta definida en la p. 575.

b) Usese esta identidad para probar que $\lim_{s\to 1+}(s-1)\zeta(s)=1$.

11. *Criterio integral de convergencia*

a) Sea $f(x)$ una función positiva y decreciente para $x \geqslant 1$. Demostrar que la integral impropia $\int_1^{\infty} f(x)\,dx$ y la serie infinita $\sum_{k=1}^{\infty} f(k)$ convergen o divergen simultáneamente.

b) Demostrar que en cualquiera de esos casos el límite

$$\lim_{n\to\infty}\left(\int^{n} f(x)\,dx-\sum_{k=1}^{n} f(k)\right)$$

existe.

c) Aplíquese este criterio para probar que la serie

$$\sum_{n=2}^{\infty}\frac{1}{n\log^{\alpha} n}$$

converge cuando $\alpha > 1$ y diverge cuando $\alpha \leq 1$.

8

Series trigonométricas

Las funciones representadas por series de potencias o, como Lagrange las llamó, las *"funciones analíticas"*, desempeñan sin duda un papel muy importante en el Análisis. No obstante, la clase de funciones analíticas es demasiado restringida en muchos casos. Fue por ello de suma importancia para la Matemática, y para una gran variedad de aplicaciones, el que Fourier, en su "Théorie analytique de la chaleur",[1] observara e ilustrara mediante gran número de ejemplos el hecho de que *series trigonométricas* convergentes de la forma

$$(1) \qquad f(x) = \frac{a_0}{2} + \sum_{\nu=1}^{\infty} (a_\nu \cos \nu x + b_\nu \operatorname{sen} \nu x),$$

con coeficientes constantes a_ν, b_ν, fuesen capaces de representar una amplia clase de funciones $f(x)$ "arbitrarias", clase que incluye esencialmente todas las funciones de interés específico, ya sea que estén definidas geométricamente por medios mecánicos o de alguna otra manera. Más aún, es posible expresar en esa forma funciones que tienen discontinuidades de salto o que obedecen leyes de formación diferentes en diferentes intervalos.

Poco después del dramático descubrimiento de Fourier, las "series de Fourier" fueron reconocidas, no sólo como una herramienta de gran utilidad en la Física y en la Mecánica, sino también como una de las fuentes más fructíferas de gran número de bellos resultados puramente matemáticos. Cauchy y, muy especialmente, Dirichlet, entre los años 1820 y 1830 pusieron sobre bases sólidas el razonamiento algo heurístico e incompleto de Fourier, haciendo el tema tan accesible como su importancia lo requería.

A pesar de lo "arbitrario" de las funciones expresables mediante series trigonométricas, aquéllas están esencialmente sujetas a satisfacer la condición de periodicidad con período 2π, puesto que cada término de la serie tiene

[1] Puede verse también, por ejemplo, *The Analytical Theory of Heat*, de Joseph Fourier, versión en inglés, Dover Publications, 1955.

este período. Sin embargo, como veremos después, esta restricción no es esencial, siempre que se considere una función restringida a un intervalo finito a partir del cual pueda hacerse fácilmente la extensión como una función periódica.

Este capítulo constituye una introducción elemental a la teoría de las series de Fourier, dejándose a un lado refinamientos más avanzados.

Después de una discusión preliminar de las funciones periódicas, demostraremos el teorema principal, en el cual se establece la validez del desarrollo trigonométrico para una amplia clase de funciones.

En las secciones subsiguientes discutiremos tópicos algo más avanzados, tales como la convergencia uniforme y absoluta de las series de Fourier y la aproximación mediante polinomios de funciones continuas arbitrarias. En el Apéndice discutiremos la teoría de los polinomios de Bernoulli y sus aplicaciones.

8.1 Funciones periódicas

a. Comentarios generales. Extensión periódica de una función

Las funciones sen nx y cos nx son funciones periódicas de x con período común 2π. Así también, cualquier suma finita, o infinita pero convergente, del tipo (1) es también una función periódica con período 2π. Haremos ahora algunas observaciones generales relativas a las funciones periódicas, ampliando aquéllas que se hicieron en el capítulo 4, p. 355.

La periodicidad de una función $f(x)$ con período T se expresa mediante la ecuación

$$f(x+T) = f(x), \tag{2a}$$

válida para todos los valores de x.[1] El tener período T implica que $f(x)$ tiene también los períodos $\pm T$, $\pm 2T$, . . . , $\pm mT$, . . . , y

$$f(x \pm mT) = f(x) \tag{2b}$$

para todos los enteros m.

En ciertos casos especiales puede ocurrir también que $f(x)$ tenga un período más corto. Por ejemplo, la función sen $(4\pi x/T)$ tiene el período T así como también el período más pequeño $T/2$.

[1] Al representar funciones periódicas, es a menudo conveniente considerar la variable independiente x como un punto sobre una circunferencia, en vez de hacerlo sobre una recta. Para una función $f(x)$ con período 2π, consideramos x como el ángulo en el centro de un círculo de radio unitario, medido entre un radio inicial arbitrario y el radio correspondiente a un punto variable sobre la circunferencia. En esa forma la periodicidad de $f(x)$ significa que a cada punto de la circunferencia le corresponde uno y sólo un valor de la función, aunque el ángulo x mismo está determinado solamente a menos de múltiplos de 2π.

Como vimos en el capítulo 4, p. 355, una función $f(x)$ definida en un intervalo cerrado $a \leq x \leq b$ puede ser desarrollada como una función periódica con período $T = b - a$ para todos los valores de x, definiendo la función en intervalos adyacentes sucesivos de longitud T fuera del intervalo original $a \leq x \leq b$ por la relación de periodicidad

$$(2c) \qquad f(x + nT) = f(x), \qquad n = \pm 1, \pm 2, \ldots .$$

La función extendida no está definida en forma única ni es necesariamente continua en los puntos extremos $x = a + nT = b + (n - 1)T$ de los intervalos de longitud T. Podemos admitir funciones $f(x)$ con discontinuidades de salto en los puntos $x = \xi$, que sean continuas a uno y otro lado de ξ pero no necesariamente definidas o continuas en el punto ξ mismo.

Establecido lo anterior, damos las siguientes notaciones y la definición de $f(\xi)$, que serán de utilidad en todo este capítulo: denotamos los límites derecho e izquierdo de $f(x)$ en $x = \xi$, respectivamente, por

$$(3a) \qquad f(\xi + 0) = \lim_{\varepsilon \to 0} f(\xi + \varepsilon^2),$$

$$(3b) \qquad f(\xi - 0) = \lim_{\varepsilon \to 0} f(\xi - \varepsilon^2);$$

Es conveniente asignar a $f(\xi)$, el valor de f en el punto de discontinuidad, el *valor medio*

$$(4) \qquad f(\xi) = \tfrac{1}{2}[f(\xi + 0) + f(\xi - 0)]$$

Con esta convención no quedan ya restricciones para extender periódicamente la función original, del intervalo cerrado $a \leq x \leq b$ a todos los valores de x, aun en los casos en que $f(a) \neq f(b)$. Sólo se necesita tener cuidado con los valores que $f(x)$ toma en los puntos de discontinuidad de salto, en particular, si los valores definidos originalmente para $f(a)$ y $f(b)$ no coinciden. Para definir la extensión periódica tiene que usarse el valor medio $\frac{1}{2}[f(a) + f(b)]$ en lugar de los valores $f(a)$ y $f(b)$.

b. Integrales calculadas sobre un período

La gráfica de una función periódica tiene claramente la misma forma en dos intervalos consecutivos cualesquiera correspondientes a un período. Esto tiene una consecuencia importante: para una función periódica $f(x)$ de período T y para una constante arbitraria a,

$$(5) \qquad \int_{-a}^{T-a} f(x)\, dx = \int_{0}^{T} f(x)\, dx,$$

o, en palabras: la integral de una función periódica evaluada sobre un intervalo "período" de longitud T tiene siempre el mismo valor, independientemente de la localización del intervalo.

Para demostrar esto sólo se necesita observar que, en virtud de la ecuación $f(\xi - T) = f(\xi)$, la substitución $x = \xi - T$, para α y β cualesquiera, conduce a

$$\int_{\alpha}^{\beta} f(x)\,dx = \int_{\alpha+T}^{\beta+T} f(\xi)\,d\xi = \int_{\alpha+T}^{\beta+T} f(x)\,dx.$$

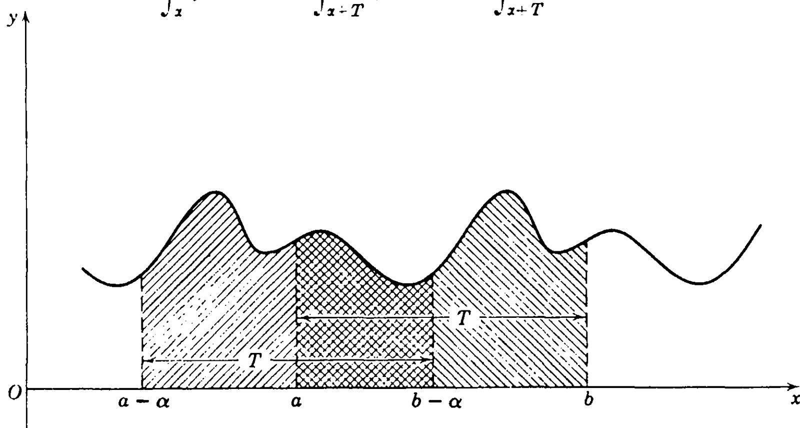


Figura 8.1 Se ilustra la integral calculada sobre un período completo.

En particular, para $\alpha = -a$ y $\beta = 0$ se tiene

$$\int_{-a}^{0} f(x)\,dx = \int_{T-a}^{T} f(x)\,dx$$

y, por tanto,

$$\begin{aligned}\int_{-a}^{T-a} f(x)\,dx &= \int_{-a}^{0} f(x)\,dx + \int_{0}^{T-a} f(x)\,dx \\ &= \int_{T-a}^{T} f(x)\,dx + \int_{0}^{T-a} f(x)\,dx \\ &= \int_{0}^{T} f(x)\,dx,\end{aligned}$$

como se afirmó. Este resultado se comprende fácilmente por medio de la figura 8.1. Basta recordar el significado geométrico de la integral.

c. Vibraciones armónicas

Las funciones periódicas más simples, a partir de las cuales construiremos las más generales, son las funciones $a \operatorname{sen} \omega x$ y $a \cos \omega x$, o bien, con

mayor generalidad, $a \operatorname{sen} \omega(x - \xi)$ y $a \cos \omega(x - \xi)$, en donde $a\ (\geq 0)$, $\omega\ (> 0)$ y ξ son constantes. Estas funciones representan *"vibraciones sinusoidales"* o *vibraciones (u oscilaciones) armónicas simples.*[1] El *período de vibración* es $T = 2\pi/\omega$. El número ω se llama la *frecuencia angular* o *circular* de las vibraciones.[2] Puesto que $1/T = \omega/2\pi$ es el número de vibraciones en la unidad de tiempo, esto es, la frecuencia, ω es el número de vibraciones en el tiempo 2π. El número a se llama la *amplitud* de la vibración; representa el máximo valor de la función $a \operatorname{sen} \omega(x - \xi)$ o bien de $a \cos \omega(x - \xi)$, pues tanto el seno como el coseno tienen valor máximo igual a 1. Los números $\omega(x - \xi)$ y $\omega\xi$ se llaman, respectivamente, la *fase* y el desplazamiento o *corrimiento de fase.*

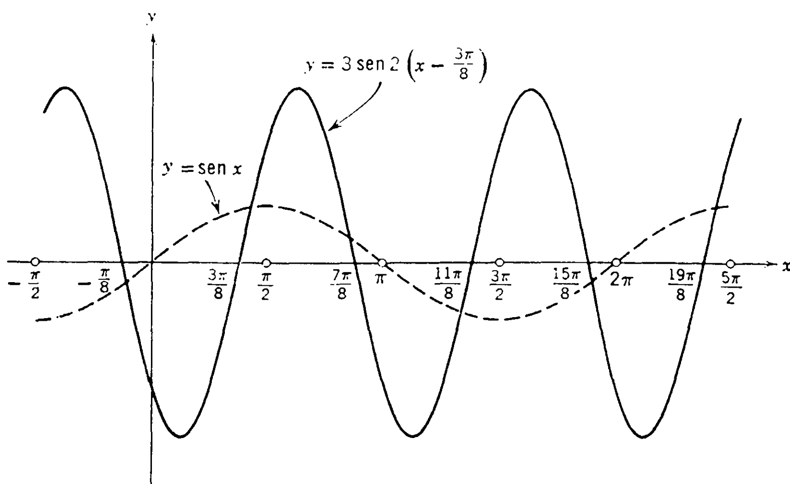


Figura 8.2 Vibraciones sinusoidales.

Las funciones $a \operatorname{sen} \omega(x - \xi)$ se obtienen gráficamente "estirando" la curva *seno* en las proporciones $1 : \omega$ a lo largo del eje x y $a : 1$ a lo largo del eje y, y trasladando después la curva una distancia ξ en el sentido positivo del eje x (ver Fig. 8.2).

Mediante las fórmulas de adición para funciones trigonométricas es también posible representar las vibraciones armónicas anteriores por $\alpha \cos \omega x + \beta \operatorname{sen} \omega x$ y $\beta \cos \omega x - \alpha \operatorname{sen} \omega x$, respectivamente, siendo $\alpha = -a \operatorname{sen} \omega\xi$ y $\beta = a \cos \omega\xi$. Recíprocamente, toda función de la forma $\alpha \cos \omega x + \beta \operatorname{sen} \omega x$ representa una vibración sinusoidal $a \operatorname{sen} \omega(x - \xi)$ de amplitud $a = \sqrt{\alpha^2 + \beta^2}$

---

[1] Cualquiera de estas dos fórmulas representa por sí misma (para todos los valores de a y ξ) el conjunto de todas las vibraciones sinusoidales. Las dos fórmulas son equivalentes, pues $a \operatorname{sen} \omega(x - \xi) = a \cos \omega \left[x - \left(\xi + \frac{\pi}{2\omega} \right) \right]$.

[2] Obsérvese que distinguimos entre *"frecuencia"* y *"frecuencia angular o circular".*

y *corrimiento de fase* $\omega\xi$ dados por las ecuaciones $\alpha = -a \operatorname{sen} \omega\xi$, y $\beta = a \cos \omega\xi$. Usando la expresión $\alpha \cos \omega x + \beta \operatorname{sen} \omega x$ podemos escribir inmediatamente la suma de dos o más de tales funciones que tengan la misma frecuencia angular ω en la forma de otra vibración con frecuencia angular ω.

Como vimos anteriormente, las funciones periódicas aparecen cuando se desea representar en forma paramétrica una curva cerrada. Naturalmente, tales funciones pueden usarse para representar fenómenos inducidos por un movimiento circular, por ejemplo, un proceso que se repite periódicamente en sincronía con un volante. Además, están asociadas a todos los fenómenos de vibración.

8.2 Superposición de vibraciones armónicas

a. Armónicas. Polinomios trigonométricos

Aunque muchas vibraciones son puramente sinusoidales (ver p. 422), la mayoría de los movimientos periódicos tienen un carácter más complejo, siendo obtenidos por "superposición" de varias vibraciones sinusoidales. Matemáticamente, el movimiento de un punto sobre una recta, cuya coordenada x es una función del tiempo, puede representarse mediante una función que sea la suma de cierto número de funciones periódicas puras del tipo considerado antes. Las componentes armónicas de la función se superponen (esto es, sus ordenadas se suman). En esta superposición se considera que las frecuencias angulares (y, claro está, los períodos) de las vibraciones superpuestas son todas diferentes, pues la superposición de dos vibraciones sinusoidales de la misma frecuencia angular conduce a otra vibración sinusoidal de igual frecuencia, como se demostró con anterioridad.

Para la superposición de dos vibraciones sinusoidales con diferentes frecuencias angulares, ω_1 y ω_2, existen dos posibilidades fundamentalmente distintas, dependiendo de que ω_1/ω_2 sea o no racional, o, como también se dice, de que las frecuencias sean *conmensurables* o *inconmensurables.*

Como un ejemplo de la primera posibilidad consideremos el caso en que la segunda frecuencia angular es dos veces la primera: $\omega_2 = 2\omega_1$. El período de la segunda vibración es entonces la mitad del período de la primera, $2\pi/2\omega_1 = T_2 = T_1/2$, de manera que aquélla no sólo tiene el período T_2 sino que también tiene el período doble, T_1, pues la función se repite después de este doble período. La función formada por superposición deberá, de la misma manera, tener el período T_1. La segunda vibración, cuya frecuencia angular es doble y cuyo período es la mitad del primero, se llama una *primera armónica* de la primera vibración (la fundamental).

Asertos correspondientes son válidos si se introduce una tercera vibración de frecuencia angular $\omega_3 = 3\omega_1$. Aquí nuevamente la función sen $3\omega_1 x$ necesariamente se repite con período $2\pi/\omega_1 = T_1$. Esta vibración se llama

una *segunda armónica* de la vibración dada. Análogamente, podemos considerar las armónicas tercera, cuarta, ..., $(n-1)$-ésima, con frecuencias angulares $\omega_4 = 4\omega_1$, $\omega_5 = 5\omega_1$, $\ldots, \omega_n = n\omega_1$, y corrimientos de fase cualesquiera. Cualquiera de tales armónicas necesariamente se repite después del período $T_1 = 2\pi/\omega_1$ y, en consecuencia, toda función obtenida por superposición de cierto número de vibraciones, armónicas de una fundamental dada de frecuencia angular ω_1, es a su vez una función periódica con período $2\pi/\omega_1 = T_1$. Superponiendo vibraciones cuyas frecuencias angulares varíen desde la de la fundamental hasta la de la $(n-1)$-ésima armónica, se obtiene una función periódica en la forma de un ***polinomio trigonométrico***,

$$S_n(x) = \frac{a_0}{2} + \sum_{\nu=1}^{n} (a_\nu \cos \nu\omega x + b_\nu \operatorname{sen} \nu\omega x). \tag{6}$$

(La constante $a_0/2$, que no afecta la periodicidad, se deja indeterminada por las razones de conveniencia que veremos después.) Como esta función contiene $2n + 1$ constantes arbitrarias, a_ν, b_ν, estamos en condiciones de generar curvas que en nada se parezcan a las sinusoides originales. Las figuras 8.3 a 8.5 constituyen ilustraciones gráficas.

El término "armónica" hace alusión a la Acústica,[1] en donde una vibración fundamental de frecuencia angular ω corresponde a un tono de cierta altura, y las armónicas primera, segunda, tercera, etc., corresponden a la sucesión de armónicas de la fundamental, es decir, a la primera octava, a la segunda octava, etc.

En general, cuando se superponen vibraciones cuyas frecuencias angulares guardan relaciones racionales entre sí, es posible representar dichas frecuencias como múltiplos enteros de una frecuencia fundamental común.

La superposición de dos vibraciones de frecuencias angulares ω_1 y ω_2 inconmensurables representa, sin embargo, un fenómeno diferente. En este caso la superposición de vibraciones sinusoidales ya no es periódica. Sin entrar en una discusión detallada, diremos que tales funciones tienen un carácter "aproximadamente periódico" o, como a veces se dice, son *cuasi periódicas*.

*b. Batidos

En un comentario final acerca de la superposición de vibraciones sinusoidales, nos referiremos al fenómeno de los llamados *batidos*. Si se superponen dos vibraciones de amplitud unitaria pero de frecuencias angulares ω_1 y ω_2, diferentes entre sí, y si, por simplicidad, se toma el mismo valor

[1] En Acústica se usa también el término *sobretono*.

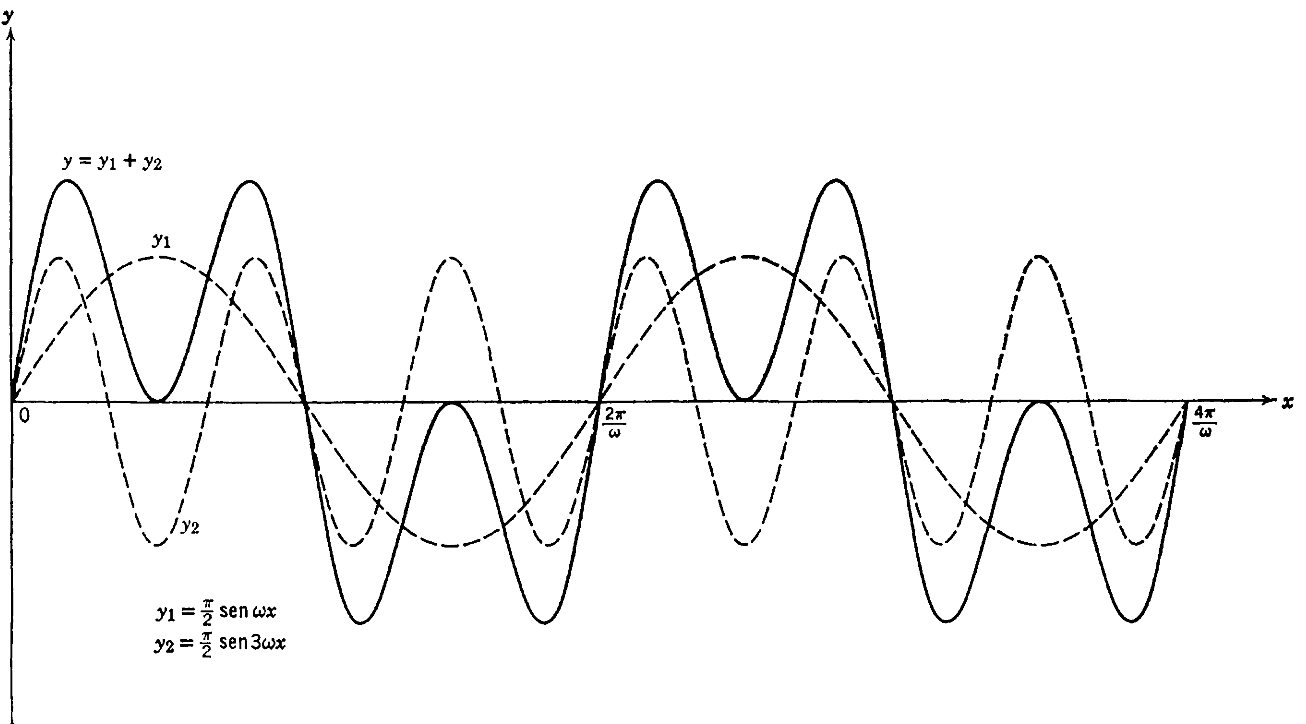


Figura 8.3 Combinación de vibraciones. (Las proporciones de la figura corresponden a la hipótesis $\omega = 1$).

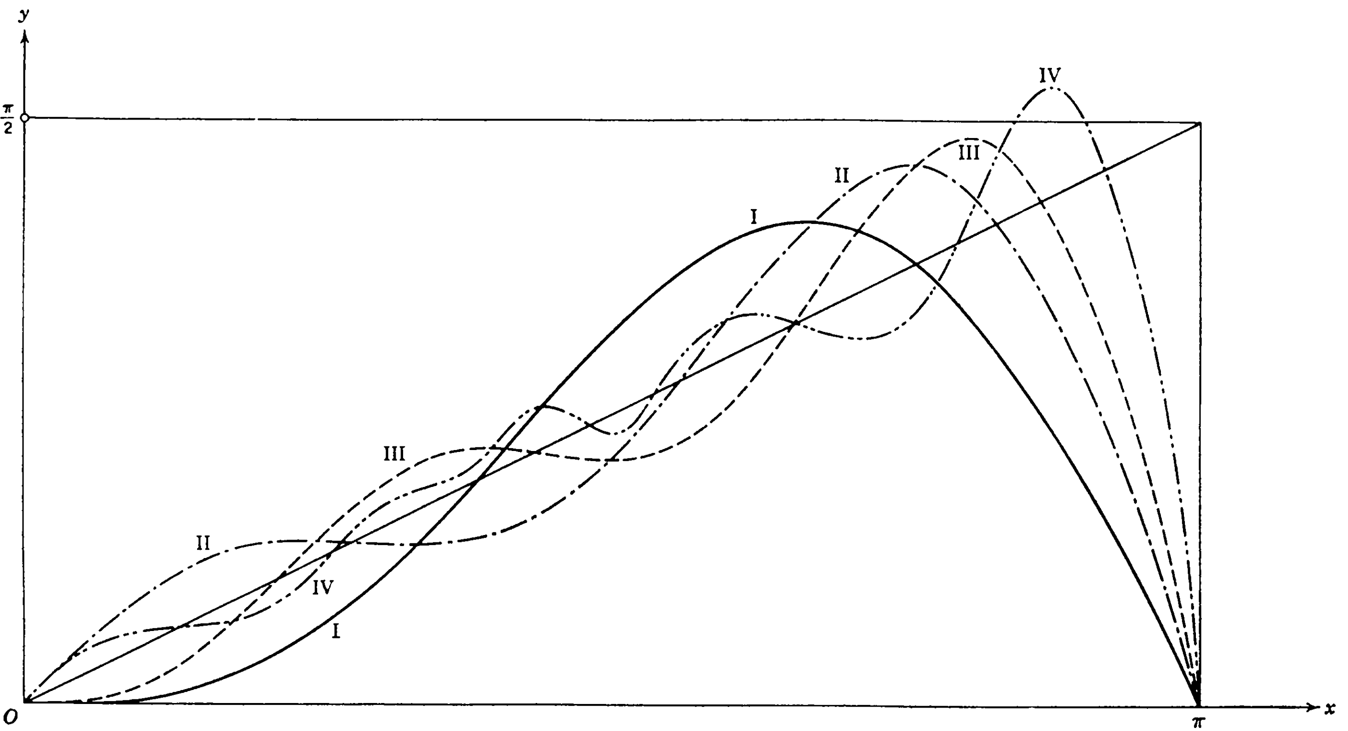


Figura 8.4 Combinación de vibraciones: I: sen x — (sen $2x$)/2; II: sen x — (sen $2x$)/2 + (sen $3x$)/3; III: sen x — (sen $2x$)/2 + (sen $3x$)/3 — (sen $4x$)/4; IV: sen x — (sen $2x$)/2 + (sen $3x$)/3 — (sen $4x$)/4 + (sen $5x$)/5 — (sen $6x$)/6 + (sen $7x$)/7.

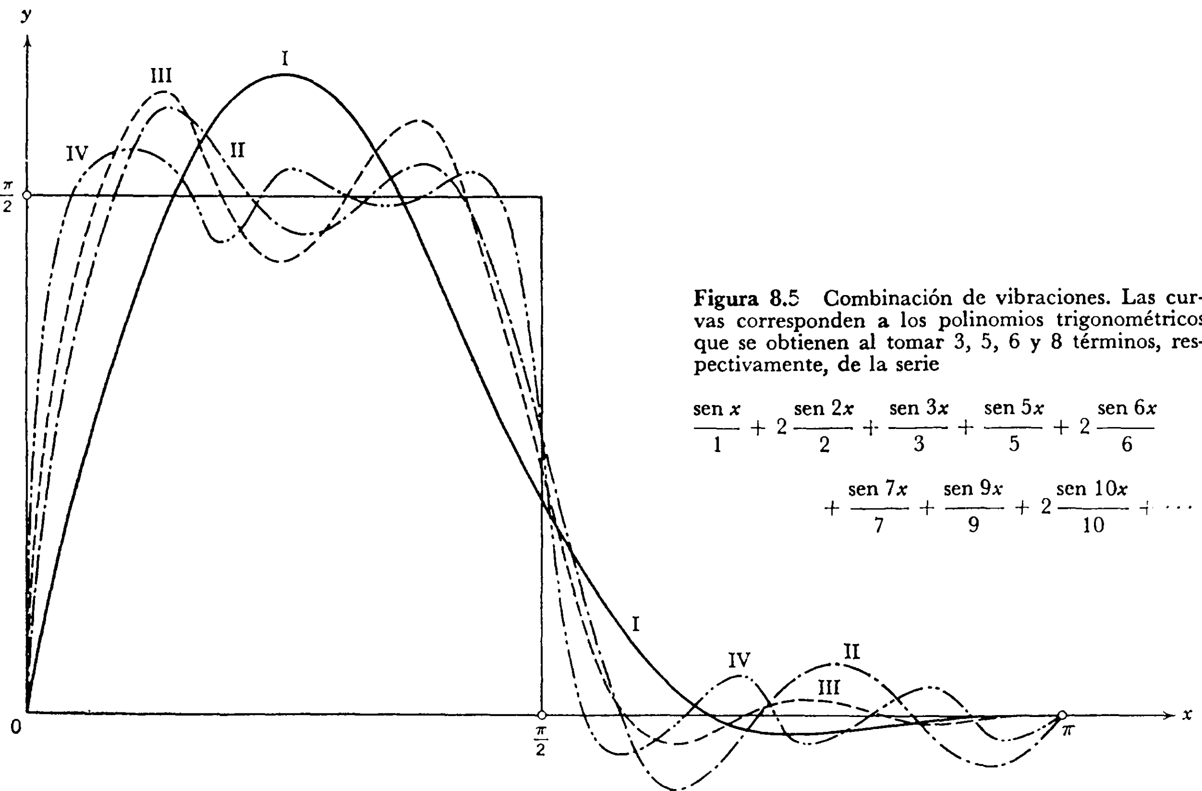


Figura 8.5 Combinación de vibraciones. Las curvas corresponden a los polinomios trigonométricos que se obtienen al tomar 3, 5, 6 y 8 términos, respectivamente, de la serie

$$\frac{\operatorname{sen} x}{1} + 2\frac{\operatorname{sen} 2x}{2} + \frac{\operatorname{sen} 3x}{3} + \frac{\operatorname{sen} 5x}{5} + 2\frac{\operatorname{sen} 6x}{6} + \frac{\operatorname{sen} 7x}{7} + \frac{\operatorname{sen} 9x}{9} + 2\frac{\operatorname{sen} 10x}{10} + \cdots$$

de ξ (ver p. 591) para ambas (la generalización al caso de fases arbitrarias se deja al lector), resulta entonces la función

$$y = \text{sen}\,\omega_1 x + \text{sen}\,\omega_2 x \qquad (\omega_1 > \omega_2 > 0).$$

Por una bien conocida fórmula trigonométrica se tiene

$$y = 2\cos\left[\tfrac{1}{2}(\omega_1 - \omega_2)x\right]\text{sen}\left[\tfrac{1}{2}(\omega_1 + \omega_2)x\right].$$

Esta ecuación representa un fenómeno que se describe como sigue: considérese una vibración de frecuencia angular $\frac{1}{2}(\omega_1 + \omega_2)$ y de período

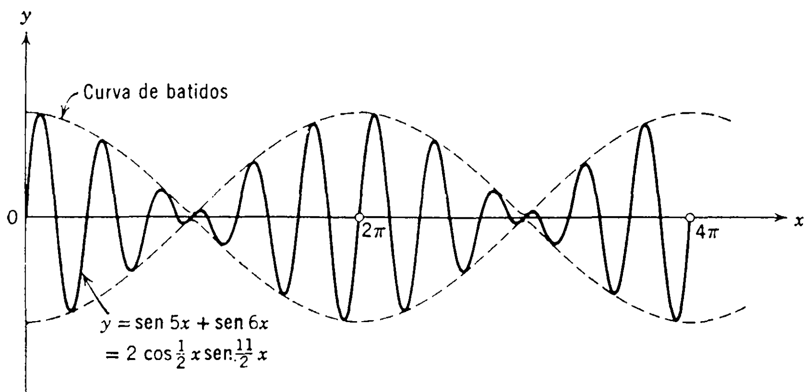


Figura 8.6 Batidos.

$4\pi/(\omega_1 + \omega_2)$. Tal vibración no tiene una amplitud constante sino una "amplitud" variable dada por la expresión

$$2\cos\left[\tfrac{1}{2}(\omega_1 - \omega_2)x\right],$$

que varía con un período mayor: $4\pi/(\omega_1 - \omega_2)$. Esta descripción es particularmente útil cuando las dos frecuencias angulares, ω_1 y ω_2, son relativamente grandes mientras que su diferencia $(\omega_1 - \omega_2)$, es relativamente pequeña. En este caso la amplitud $2\cos\left[\frac{1}{2}(\omega_1 - \omega_2)x\right]$ de la vibración, cuyo período es $4\pi/(\omega_1 + \omega_2)$, varía lentamente en comparación con el período de vibración, y este cambio de amplitud se repite en forma periódica con un período, $4\pi/(\omega_1 - \omega_2)$, grande. Tales cambios rítmicos de amplitud se denominan *batidos*. Este fenómeno se presenta con frecuencia en Acústica y en Electrónica. En radiotransmisión, las frecuencias ω_1 y ω_2 caen, por regla general, en un rango muy por encima del que puede detectar el oído humano, en tanto que la diferencia $\omega_1 - \omega_2$ cae en el rango de notas audibles. Los batidos producen entonces un tono audible aunque las vibraciones originales sean imperceptibles por el oído.

En la Fig. 8.6 se ilustra un caso en que se tienen batidos.

8.3 Notación compleja

a. Comentarios generales

Las operaciones con funciones trigonométricas se simplifican a menudo usando números complejos de acuerdo con la relación de Euler:

$$\cos\theta + i \operatorname{sen}\theta = e^{i\theta},$$

o sea,

$$\cos\theta = \tfrac{1}{2}(e^{i\theta} + e^{-i\theta}), \tag{7a}$$

$$\operatorname{sen}\theta = \frac{1}{2i}(e^{i\theta} - e^{-i\theta}). \tag{7b}$$

(Compárese con capítulo 7, p. 566.) Según esto, es posible representar las vibraciones sinusoidales en términos de las cantidades complejas $e^{i\omega x}$, $e^{-i\omega x}$, o $ae^{i\omega(x-\xi)}$, $ae^{-i\omega(x-\xi)}$, siendo a, ω y $\omega\xi$ la amplitud, la frecuencia angular y el corrimiento de fase, respectivamente. Como cabría esperar, las vibraciones reales se obtienen a partir de las expresiones complejas simplemente separando partes real e imaginaria.

Una de las ventajas de la notación compleja consiste en que las derivadas con respecto al tiempo se obtienen derivando la función exponencial compleja como si i fuese una constante real. La fórmula

$$\begin{aligned}\frac{d}{dx}a[\cos\omega(x-\xi) + i\operatorname{sen}\omega(x-\xi)]&\\ &= a\omega[-\operatorname{sen}\omega(x-\xi) + i\cos\omega(x-\xi)]\\ &= ia\omega[\cos\omega(x-\xi) + i\operatorname{sen}\omega(x-\xi)],\end{aligned}$$

que se obtiene a partir de las fórmulas para las derivadas de las funciones seno y coseno, puede escribirse en la forma concisa

$$\frac{d}{dx}ae^{i\omega(x-\xi)} = ia\omega e^{i\omega(x-\xi)}. \tag{8}$$

La integral de una función compleja $\gamma(x)$, digamos $\gamma(x) = p(x) + iq(x)$, está definida en forma natural por

$$\int \gamma(x)\,dx = \int p(x)\,dx + i\int q(x)\,dx.$$

Según esto, para $n \neq 0$ se tiene

$$\begin{aligned}\int e^{inx}\,dx &= \int \cos nx\,dx + i\int \operatorname{sen} nx\,dx\\ &= \frac{1}{n}\operatorname{sen} nx - \frac{i}{n}\cos nx = \frac{1}{in}e^{inx}.\end{aligned}$$

En particular, para cualquier entero n resulta

$$\int_{-\pi}^{\pi} e^{inx}\,dx = \begin{cases} 0 & \text{para } n \neq 0 \\ 2\pi & \text{para } n = 0. \end{cases}$$

Más generalmente, teniendo en cuenta que $e^{inx}e^{-imx} = e^{i(n-m)x}$, se obtiene para m y n enteros cualesquiera

$$\int_{-\pi}^{\pi} e^{inx}e^{-imx}\,dx = \begin{cases} 0, & \text{para } n \neq m \\ 2\pi & \text{para } n = m. \end{cases} \tag{9}$$

Estas relaciones son simplemente expresiones condensadas de las relaciones de ortogonalidad entre funciones trigonométricas (ver p. 295).

*b. Aplicación al caso de corrientes alternas

Ilustramos aquí las ideas anteriores mediante un ejemplo importante. Denotaremos la variable independiente, el tiempo, con t, en vez de hacerlo con x.

Consideremos un circuito eléctrico con resistencia R e inductancia L, al cual se aplica una "fuerza electromotriz" (voltaje) externa E. Si se trata de corriente "continua" o directa, el voltaje E es constante y la corriente I está dada por la ley de Ohm: $E = RI$. En el caso de corriente alterna, sin embargo, E (y, por lo tanto, I) es una función del tiempo t, y la *ley de Ohm* toma la *forma generalizada* (ver p. 653)

$$E - L\frac{dI}{dt} = RI. \tag{10}$$

Supongamos que la fuerza electromotriz externa, E, es sinusoidal, con frecuencia angular ω, y está dada por $\varepsilon\cos\omega t$, o bien por $\varepsilon\,\text{sen}\,\omega t$, y combinemos formalmente ambas posibilidades en la forma compleja

$$E = \varepsilon e^{i\omega t} = \varepsilon\cos\omega t + i\varepsilon\,\text{sen}\,\omega t,$$

donde ε representa la amplitud. Es a menudo útil admitir valores *complejos* también para la *amplitud:*

$$\varepsilon = |\varepsilon|\,e^{-i\eta}.$$

Entonces,

$$E = |\varepsilon|\,e^{i(\omega t-\eta)} = |\varepsilon|\,\{\cos(\omega t - \eta) + i\,\text{sen}\,(\omega t - \eta)\}.$$

Podemos operar con este *"voltaje complejo"* E y la correspondiente *"corriente compleja"* I como si i fuese un parámetro real. El significado de la relación compleja entre las cantidades complejas E e I es entonces que la corriente que corresponde a una fuerza electromotriz $\varepsilon\cos\omega t$ es la parte real de I, en tanto que la correspondiente a una fuerza electromotriz $\varepsilon\,\text{sen}\,\omega t$ es la parte imaginaria de I. La corriente compleja está dada por una expresión de la forma

$$I = \alpha e^{i\omega t} = \alpha(\cos\omega t + i\,\text{sen}\,\omega t),$$

que también es sinusoidal con frecuencia angular ω. La derivada de I está dada entonces formalmente por

$$\frac{dI}{dt} = i\alpha\omega e^{i\omega t}$$

$$= \alpha\omega(-\text{sen}\,\omega t + i\cos\omega t) = i\omega I.$$

Substituyendo estas cantidades en la forma *generalizada* de la *ley de Ohm* (Ec. 10), y dividiendo por $e^{i\omega t}$, se obtiene la ecuación

$$\varepsilon - \alpha Li\omega = R\alpha,$$

o sea,

$$\alpha = \frac{\varepsilon}{R + i\omega L},$$

así como también

$$E = (R + i\omega L)I = WI.$$

Podemos considerar esta última ecuación como la *ley de Ohm* en forma compleja para corrientes alternas, llamando a la cantidad

$$W = R + i\omega L$$

resistencia compleja del circuito. La ley de Ohm tiene entonces la misma forma para corrientes "continuas": la corriente es igual al voltaje dividido por la resistencia.

Escribiendo la resistencia compleja W en la forma

$$W = we^{i\delta} = w\cos\delta + iw \operatorname{sen}\delta,$$

donde

$$|W| = w = \sqrt{(R^2 + L^2\omega^2)}, \qquad \tan\delta = \frac{\omega L}{R},$$

obtenemos

$$I = \frac{\varepsilon}{w} e^{i(\omega t - \delta)} = \frac{E}{W}.$$

De acuerdo con esta fórmula, la corriente tiene el mismo período (y la misma frecuencia angular) que el voltaje. La amplitud, α, de la corriente está relacionada con la amplitud, ε, de la fuerza electromotriz, para ε real, mediante la expresión

$$\alpha = \frac{\varepsilon}{w}$$

Además, existe una diferencia de fase entre la corriente y el voltaje. La corriente no alcanza su máximo al mismo tiempo que el voltaje sino un tiempo δ/ω más tarde y, por supuesto, lo mismo es cierto para el mínimo. En ingeniería eléctrica la cantidad $w = \sqrt{R^2 + L^2\omega^2}$ se llama a menudo la *impedancia* o *resistencia en corriente alterna* del circuito, para la frecuencia angular ω. El corrimiento de fase, dado por lo general en grados, se llama algunas veces *defasaje.*

Si la amplitud ε es compleja y tiene la forma

$$\varepsilon = |\varepsilon|\, e^{-i\eta},$$

nada esencial cambia en la forma de la ley de Ohm, excepto que η es un corrimiento de fase adicional y se tiene entonces

$$E = |\varepsilon|\, e^{i(\omega t - \eta)},$$

$$I = \frac{E}{W} = \frac{|\varepsilon|}{|W|} e^{i\omega t} e^{-i(\delta+\eta)}.$$

c. Notación compleja para polinomios trigonométricos

Una vibración compuesta del tipo

$$(11) \qquad S_n(x) = \tfrac{1}{2}a_0 + \sum_{\nu=1}^{n} (a_\nu \cos \nu x + b_\nu \operatorname{sen} \nu x)$$

(donde, por brevedad, se ha tomado $\omega = 1$) puede reducirse a la forma compleja mediante las substituciones

$$\cos \nu x = \tfrac{1}{2}(e^{i\nu x} + e^{-i\nu x}) \qquad \text{y} \qquad \operatorname{sen} \nu x = -\tfrac{1}{2}i(e^{i\nu x} - e^{-i\nu x}).$$

La primera expresión toma entonces la forma más simple

$$(12) \qquad S_n(x) = \sum_{\nu=-n}^{n} \alpha_\nu e^{i\nu x},$$

en la que los números complejos α_ν están relacionados con los números reales a_0, a_ν y b_ν por medio de las ecuaciones

$$(13a) \qquad \begin{cases} \alpha_\nu = \tfrac{1}{2}(a_\nu - ib_\nu), \\ \alpha_{-\nu} = \tfrac{1}{2}(a_\nu + ib_\nu), \qquad \text{para} \quad \nu = 1, 2, \ldots, n. \\ \alpha_0 = \tfrac{1}{2}a_0. \end{cases}$$

Al resolver estas ecuaciones para los a_ν y b_ν, se encuentra que

$$(13b) \qquad \begin{aligned} a_\nu &= \alpha_\nu + \alpha_{-\nu}, \\ b_\nu &= i(\alpha_\nu - \alpha_{-\nu}). \end{aligned}$$

(Queda incluido el caso $\nu = 0$.)

Recíprocamente, cualquier expresión de la forma

$$\sum_{\nu=-n}^{n} \alpha_\nu e^{i\nu x}$$

puede considerarse como una función que representa la superposición de vibraciones escritas en forma compleja. El resultado de esta superposición es real si, y sólo si, $\alpha_\nu + \alpha_{-\nu}$ es real y $\alpha_\nu - \alpha_{-\nu}$ es imaginario puro; o sea, si α_ν y $\alpha_{-\nu}$ son números complejos conjugados.

d. Una fórmula trigonométrica

Como aplicación de la notación compleja, demostraremos la identidad siguiente:

$$(14) \qquad \begin{aligned} \sigma_n(\alpha) &= \tfrac{1}{2} + \cos \alpha + \cos 2\alpha + \cdots + \cos n\alpha \\ &= \frac{\operatorname{sen}\,(n + \frac{1}{2})\alpha}{2 \operatorname{sen} \frac{1}{2}\alpha}, \end{aligned}$$

que necesitaremos más adelante en este mismo capítulo. La fórmula tiene sentido solamente cuando $\operatorname{sen} \frac{1}{2}\alpha \neq 0$, es decir, cuando α no tiene alguno de los valores $0, \pm 2\pi, \pm 4\pi, \ldots$. Sin embargo, una vez establecida la fórmula para $\operatorname{sen} \frac{1}{2}\alpha \neq 0$, es posible concluir que la expresión $[\operatorname{sen}(n + \frac{1}{2})\alpha]/(2 \operatorname{sen} \frac{1}{2}\alpha)$ es una función continua de α, para todo α, si se define su valor en los puntos excepcionales como el valor de $\sigma_n(\alpha)$, es decir, $n + \frac{1}{2}$.

Para la demostración, reemplazamos la función coseno por su expresión exponencial [ver fórmula (13*a*) con $a_\nu = 1$, $b_\nu = 0$]:

$$\sigma_n(\alpha) = \tfrac{1}{2} \sum_{\nu=-n}^{n} e^{i\nu\alpha}.$$

En el lado derecho tenemos una progresión geométrica con el cociente común $q = e^{i\alpha} = \cos\alpha + i \operatorname{sen}\alpha$. Por tanto, q puede tomar el valor 1 sólo si $\cos\alpha = 1$ y $\operatorname{sen}\alpha = 0$, o sea, si α tiene uno de los valores excepcionales $0, \pm 2\pi, \pm 4\pi, \ldots$. Cuando α tiene cualquier otro valor la fórmula ordinaria para la suma conduce a

$$\begin{aligned} \sigma_n(\alpha) &= \tfrac{1}{2} e^{-in\alpha} \frac{1 - q^{2n+1}}{1 - q} \\ &= \tfrac{1}{2} \cdot \frac{e^{-in\alpha} - e^{(n+1)i\alpha}}{1 - e^{i\alpha}}. \end{aligned}$$

Multiplicando el numerador y el denominador por $e^{-i\alpha/2}$ se obtiene

$$\sigma_n(\alpha) = \frac{\operatorname{sen}(n + \frac{1}{2})\alpha}{2 \operatorname{sen} \frac{1}{2}\alpha},$$

como se afirmó antes.

Al integrar $\sigma_n(t)$ sobre $0 \leq t \leq \pi$, se encuentra, independientemente de n, el útil resultado siguiente:

$$(15) \qquad \int_0^\pi \frac{\operatorname{sen}(n + \frac{1}{2})t}{2 \operatorname{sen} \frac{1}{2}t} dt = \int_0^\pi \left(\tfrac{1}{2} + \sum_{\nu=1}^{n} \cos \nu t \right) dt, = \tfrac{1}{2}\pi$$

pues la integral de cada término de la serie se anula.

8.4 Series de Fourier

a. Coeficientes de Fourier

Los *polinomios trigonométricos* de orden n,

$$(16) \qquad f(x) = S_n(x) = \tfrac{1}{2}a_0 + \sum_{\nu=1}^{n} (a_\nu \cos \nu x + b_\nu \operatorname{sen} \nu x)$$

dependen de los $2n + 1$ coeficientes a_ν y b_ν. Es notable que estos *"coeficientes de Fourier"* puedan ser expresados en forma sencilla en términos de los valores $f(x)$ de la suma, mediante las fórmulas

$$(17) \qquad a_\mu = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos \mu x \, dx, \; b_\mu = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \operatorname{sen} \mu x \, dx.$$

La demostración se realiza multiplicando (16) por $\cos \mu x$ o por sen μx, e integrando después. Las relaciones de ortogonalidad (ver p. 295) conducen de inmediato a estas fórmulas, puesto que sólo los términos con $\nu = \mu$ dan una contribución no nula.

En la terminología de los números complejos,

$$f(x) = S_n(x) = \sum_{\nu=-1}^{n} \alpha_\nu e^{i\nu x}, \tag{16a}$$

con $\quad a_\nu = \alpha_\nu + \alpha_{-\nu}; \qquad b_\nu = i(\alpha_\nu - \alpha_{-\nu}).$

Las expresiones correspondientes para los coeficientes complejos de Fourier son

$$\alpha_\nu = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x) e^{-i\nu x}\, dx, \tag{17a}$$

como se deduce también de las relaciones complejas de ortogonalidad [(9), p. [599].

De paso se nota que el factor $\frac{1}{2}$ en la notación para el término constante $\frac{1}{2}a_0$ de (16) sirve sólo para hacer que la fórmula (17) sea válida también para $\nu = 0$.

El teorema principal relativo a las series de Fourier surge ahora de una pregunta natural: al hacer que el grado n del polinomio de Fourier (16) tienda a infinito ¿se tiene la posibilidad de representar funciones $f(x)$ periódicas con período 2π pero, por lo demás, esencialmente arbitrarias?

El resultado de los siguientes párrafos será sin duda que: *Cualquier función periódica $f(x)$, continua a trozos y con derivadas de primero y segundo orden continuas a trozos, puede representarse mediante una "serie de Fourier" infinita,*

$$f(x) = \frac{a_0}{2} + \sum_{\nu=1}^{\infty} (a_\nu \cos \nu x + b_\nu \operatorname{sen} \nu x),$$

o bien, en notación compleja,

$$f(x) = \sum_{\nu=-\infty}^{\infty} \alpha_\nu e^{i\nu x};$$

con coeficientes dados por (17) y (17a).

b. Lema básico

En primer lugar, recordemos que, por definición, una función *continua a trozos* en un intervalo es una función continua excepto en un número finito de discontinuidades por salto dentro del intervalo.

Recordemos además que el valor de una función periódica $f(x)$ en un punto de discontinuidad se ha definido [Ec. (4), p. 589] como el promedio de los valores límite derecho e izquierdo.

Una función $f(x)$ es *continua a trozos* y tiene primera y segunda derivadas continuas a trozos si es posible dividir el intervalo total en un número finito de subintervalos, de manera tal que f, f' y f'' sean continuas en cada subintervalo abierto y se aproximen a límites definidos en los puntos extremos.

La clave para la demostración del teorema principal es un hecho muy sencillo, establecido por el siguiente Lema.

LEMA. *Si una función, $k(x)$, y su primera derivada, $k'(x)$, son continuas a trozos en el intervalo $a \leq x \leq b$, entonces la integral*

$$K_\lambda = \int_a^b k(x) \operatorname{sen} \lambda x \, dx$$

tiende a cero cuando $\lambda \rightarrow \infty$.

DEMOSTRACIÓN. Para demostrar este lema hacemos uso de la integración por partes. Supóngase que k y k' son continuas en $a \leq x \leq b$. Se tiene entonces

$$(18) \qquad K_\lambda = \int_a^b k(x) \operatorname{sen} \lambda x \, dx$$

$$= \frac{1}{\lambda}\left[k(a) \cos \lambda a - k(b) \cos \lambda b + \int_a^b k'(x) \cos \lambda x \, dx \right]$$

Es claro que cuando λ crece el lado derecho tiende a cero. Si $k(x)$ o $k'(x)$ tuviesen discontinuidades de salto (ξ) en el intervalo, dividiríamos éste en partes por medio de ξ, aplicaríamos nuestro argumento a cada parte y luego sumaríamos los resultados.

Sin demostrarlo, diremos que el lema sigue siendo válido, en realidad, sin necesidad de la hipótesis que se refiere a la existencia de la derivada $k'(x)$. La continuidad a trozos de k es suficiente. La demostración con estas condiciones más débiles se basa en el hecho de que para $\lambda \neq 0$ la función $\operatorname{sen} \lambda x$ es alternadamente positiva y negativa en intervalos consecutivos de longitud π/λ. Para valores grandes de λ las contribuciones a la integral provenientes de intervalos contiguos casi se cancelan entre sí como consecuencia de la continuidad de $k(x)$.

c. Demostración de que $\int_0^\infty \frac{\operatorname{sen} z}{z} dz = \frac{\pi}{2}$

Como una aplicación del Lema, evaluaremos la integral

$$(19) \qquad I = \int_0^\infty \frac{\operatorname{sen} z}{z} dz.$$

Esta integral impropia está definida por la relación

$$I = \lim_{M \rightarrow \infty} I_M$$

en la que

$$I_M = \int_0^M \frac{\operatorname{sen} z}{z}\, dz.$$

La convergencia de la integral impropia I, es decir, la existencia del límite de I_M para $M \to \infty$, ha sido demostrada ya en la p. 329. Dicha demostración se basó en la integración por partes y puede reproducirse aquí. Si, por ejemplo, $0 < M < N$, se tiene

$$(20) \qquad |I_N - I_M| = \left| \int_M^N \frac{\operatorname{sen} z}{z}\, dz \right|$$
$$= \left| -\frac{\cos z}{z} \Big|_M^N + \int_M^N \frac{\cos z}{z^2}\, dz \right|$$
$$\leq \frac{1}{M} + \frac{1}{N} + \int_M^N \frac{dz}{z^2} = \frac{2}{M}.$$

Puesto que I_N e I_M difieren en una cantidad arbitrariamente pequeña cuando M y N son suficientemente grandes, la existencia de $I = \lim_{M\to\infty} I_M$ queda garantizada por el criterio de convergencia de Cauchy. Además, haciendo que N tienda a infinito en (20) se encuentra una estimación para la "rapidez" con que I_M se acerca a su límite I:

$$(20a) \qquad |I - I_M| \leq \frac{2}{M}.$$

La expresión para I puede reescribirse de manera tal que I aparezca como un límite de integrales sobre un intervalo finito fijo. Sea p un número positivo arbitrario. Para $M = \lambda p$, con la substitución $z = \lambda x$, $dz = \lambda\, dx$ se obtiene

$$I_{\lambda p} = \int_0^{\lambda p} \frac{\operatorname{sen} z}{z}\, dz = \int_0^p \frac{\operatorname{sen} \lambda x}{x}\, dx.$$

Ahora bien, puesto que $\lambda p \to \infty$ para $\lambda \to \infty$ y un número positivo fijo p, resulta claro que

$$I = \lim_{\lambda\to\infty} \int_0^p \frac{\operatorname{sen} \lambda x}{x}\, dx.$$

O, en forma más precisa, de (20a) se concluye que

$$\left| I - \int_0^p \frac{\operatorname{sen} \lambda x}{x}\, dx \right| < \frac{2}{\lambda p}.$$

Así, para cualquier número positivo p las expresiones

$$\int_0^p \frac{\operatorname{sen} \lambda x}{x}\, dx$$

tienden a uno y el mismo límite, I, cuando $\lambda \to \infty$. Además, la convergencia es uniforme en p para valores de p mayores que cierto número positivo fijo, P. Sin duda, la diferencia entre la integral y el límite I es entonces menor que ε para $\lambda > 2/P\varepsilon$.

Apliquemos ahora el Lema de la p 604 a la función

$$k(x) = \frac{1}{x} - \frac{1}{2 \operatorname{sen} (x/2)}.$$

Definiendo $k(0) = 0$, la función $k(x)$ es continua y tiene una primera derivada continua para $0 \leq x < 2\pi$ (ver p. 483). Por tanto, el Lema muestra que

$$\int_0^p \operatorname{sen} \lambda x \left(\frac{1}{x} - \frac{1}{2 \operatorname{sen} (x/2)}\right) dx$$

tiende a cero para $\lambda \to \infty$, siempre que $0 \leq p < 2\pi$. Además, por (18), la convergencia es uniforme si $0 \leq p < \pi$, puesto que $|k(x)|$ y $|k'(x)|$ son acotadas en el intervalo $0 \leq x \leq \pi$. Del resultado anterior se deduce que para cualquier p en el intervalo $0 < p < 2\pi$ se tiene

$$\lim_{\lambda \to \infty} \int_0^p \frac{\operatorname{sen} \lambda x}{2 \operatorname{sen} (x/2)} dx = I,$$

y también que la convergencia es uniforme en p sin $P \leq p \leq \pi$, siendo P un número positivo fijo.

Ahora bien, esta integral ya ha sido evaluada [ver fórmula (15), p. 602] en el caso en que $p = \pi$ y $\lambda = n + \frac{1}{2}$; y el resultado es el valor $\pi/2$, independientemente de n. Haciendo que λ tienda a infinito a través de valores de la forma $\lambda = n + \frac{1}{2}$, se encuentra entonces para I el valor $\pi/2$:

(21) $$\int_0^\infty \frac{\operatorname{sen} z}{z} dz = \frac{\pi}{2}.$$

Se ha demostrado, además, que

(21*a*) $$\lim_{\lambda \to \infty} \int_0^p \frac{\operatorname{sen} \lambda x}{x} dx = \frac{\pi}{2},$$

cuya convergencia es uniforme para $P \leq p$, donde P es un número positivo fijo; y también que

(21*b*) $$\lim_{\lambda \to \infty} \int_0^p \frac{\operatorname{sen} \lambda x}{2 \operatorname{sen} (x/2)} dx = \frac{\pi}{2},$$

cuya convergencia es uniforme para $P \leq p \leq \pi$.

d. Desarrollo de Fourier para la función $\phi(x) = x$

El resultado anterior conduce directamente al desarrollo de Fourier de dos funciones periódicas lineales a trozos, $\phi(x)$ y $\chi(x)$, relacionadas entre sí, que se definen en el intervalo $-\pi < x < \pi$ por medio de las ecuaciones

$$\phi(x) = x$$

y

$$(22) \qquad \chi(x) = \begin{cases} \pi - x & \text{para } x > 0 \\ 0 & \text{para } x = 0. \\ -\pi - x & \text{para } x < 0. \end{cases}$$

(Ver Figs. 8.7 y 8.8.)

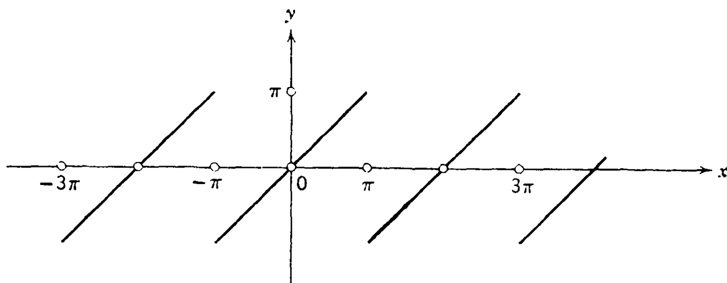


Figura 8.7 La función $\phi(x)$.

La primera función, ϕ, extendida periódicamente fuera del intervalo $-\pi < x < +\pi$, presenta discontinuidades de salto en los puntos extremos, mientras que $\chi(x)$ presenta un salto de 2π en $x = 0$. Es claro que las dos funciones extendidas periódicamente están relacionadas entre sí mediante la ecuación

$$\chi(x) = \phi(\pi - x).$$

El desarrollo de Fourier para $\chi(x)$, cuando $0 < x \leq \pi$, se deduce inmediatamente de las fórmulas (14) de p. 601 y (21*b*) de p. 606, con $\lambda = n + \frac{1}{2}$ y $p = x$, bastando para ello pasar al límite con $n \to \infty$. Resulta así la serie de Fourier

$$(23a) \qquad \chi(x) = 2\left(\operatorname{sen} x + \frac{\operatorname{sen} 2x}{2} + \frac{\operatorname{sen} 3x}{3} + \cdots\right).$$

Lo mismo es válido también para $-\pi \leq x < 0$, pues ambos lados son funciones impares de x. La serie es uniformemente convergente para $\varepsilon < |x| \leq \pi$, donde ε es un número positivo arbitrariamente pequeño. Para $x = 0$ todos

los términos de la serie son nulos y, por consiguiente, también lo es la suma, cosa que concuerda con la definición de $\chi(0)$. La identidad (23*a*) es válida para todo x por tener ambos lados el período 2π.

Puede confirmarse fácilmente que los coeficientes del desarrollo son efectivamente los coeficientes de Fourier definidos por la fórmula (17), p. 602.

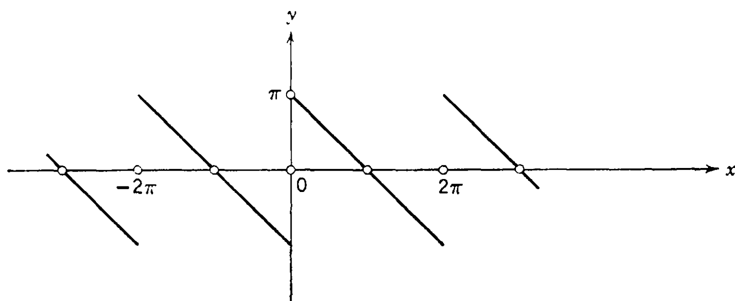


Figura 8.8 La función $\chi(x)$.

El desarrollo de Fourier para $\phi(x)$ puede obtenerse ahora directamente de la relación $\phi(x) = \chi(\pi - x)$:

$$(23b) \qquad \phi(x) = 2\sum_{\nu=1}^{\infty} (-1)^{\nu+1} \frac{\operatorname{sen} \nu x}{\nu}$$

$$= 2(\operatorname{sen} x - \tfrac{1}{2} \operatorname{sen} 2x + \tfrac{1}{3} \operatorname{sen} 3x - + \cdots).$$

En este caso la convergencia es uniforme siempre que el punto x se mantenga separado de los puntos de discontinuidad $x = \pm\pi$, mediante la condición $|x| < \pi - \varepsilon$.

Para $x = \pi/2$ se obtiene nuevamente la serie de Leibnitz:

$$\frac{\pi}{2} = 2(1 - \tfrac{1}{3} + \tfrac{1}{5} - \cdots + -).$$

Debe mencionarse que las series para χ y ϕ no convergen absolutamente. En efecto, los valores absolutos para $x = \pi/2$ forman la serie

$$2\sum_{1}^{\infty} \frac{1}{2\nu - 1},$$

que es divergente.

La fórmula (23*b*) es notable como el ejemplo de una serie infinita de funciones continuas que converge para todo x pero que tiene por suma una función discontinua, a saber, la función lineal a trozos $\phi(x)$. Cualquier suma parcial de la serie es continua, puesto que la suma de cualquier número finito de funciones continuas es a su vez continua. Debido a que una serie

infinita uniformemente convergente de funciones continuas tiene una suma continua, la serie de Fourier no puede ser uniformemente convergente en la vecindad de un punto x en el cual ϕ es discontinua, es decir, para $x = \pm\pi, \pm 3\pi, \ldots$. La Fig. 8.4, p. 595. muestra cómo las sumas parciales sucesivas, que son polinomios trigonométricos (funciones continuas), aproximan uniformemente, en un intervalo de continuidad, a la función lineal por trozos $\frac{1}{2}\phi(x)$, pero que cerca del punto extremo las funciones cambian con rapidez cada vez mayor.

e. El teorema principal sobre desarrollos de Fourier

Los coeficientes de Fourier. Después de los preparativos anteriores, se puede fácilmente asegurar la posibilidad de desarrollar en serie de Fourier una amplia clase de funciones. La forma de tal desarrollo para una función $f(x)$ con período 2π es

$$f(x) = \tfrac{1}{2}a_0 + \sum_{\nu=1}^{\infty} (a_\nu \cos \nu x + b_\nu \operatorname{sen} \nu x), \tag{24a}$$

o bien, en notación compleja,

$$f(x) = \sum_{\nu=-\infty}^{\infty} \alpha_\nu e^{i\nu x}. \tag{24b}$$

Supondremos, para comenzar, que se dispone de desarrollos uniformemente convergentes de las formas (24*a*) o (24*b*) para la función $f(x)$. Es posible entonces determinar los coeficientes a_ν, b_ν y α_ν de esos desarrollos multiplicando el primero por $\cos \mu x$, o bien $\operatorname{sen} \mu x$ y el segundo por $e^{-i\mu x}$ e integrando desde $-\pi$ hasta π, haciendo uso de las relaciones de ortogonalidad (ver pp. 295 y 599):

$$\int_{-\pi}^{\pi} \operatorname{sen} \nu x \operatorname{sen} \mu x \, dx = \int_{-\pi}^{\pi} \cos \nu x \cos \mu x \, dx = \begin{cases} 0 & \text{si } \mu \neq \nu \\ \pi & \text{si } \mu = \nu \neq 0, \end{cases}$$

$$\int_{-\pi}^{\pi} \operatorname{sen} \nu x \cos \mu x \, dx = 0, \tag{25}$$

$$\int_{-\pi}^{\pi} e^{i\nu x} e^{-i\mu x} \, dx = \begin{cases} 0 & \text{si } \mu \neq \nu \\ 2\pi & \text{si } \mu = \nu. \end{cases}$$

Representando por t la variante de integración, se obtienen de inmediato las fórmulas

$$a_\mu = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t) \cos \mu t \, dt, \qquad b_\mu = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t) \operatorname{sen} \mu t \, dt \tag{26a}$$

para $\mu = 0, 1, 2, \ldots$, y

$$\alpha_\mu = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t) e^{-i\mu x} \, dt \tag{26b}$$

para $\mu = 0, \pm 1, \pm 2, \ldots$.

Así, si $f(x)$ puede desarrollarse en una serie, ya sea de la forma (24a) o (24b), uniformemente convergente, entonces los coeficientes sólo pueden tener los valores determinados por las fórmulas (26a) y (26b). Pero, aún sin una justificación rigurosa de este procedimiento, es claro que las fórmulos (26a) o (26b) definen una sucesión de números, a_ν, b_ν y α_ν, llamados los *coeficientes de Fourier,* para toda función $f(x)$ que sea continua o continua por trozos en el intervalo $-\pi \leq x \leq \pi$.

Dada una función $f(x)$, formamos con los coeficientes definidos por (26a, b) las *sumas parciales de Fourier*

$$S_n(x) = \tfrac{1}{2}a_0 + \sum_{\nu=1}^{n} (a_\nu \cos \nu x + b_\nu \operatorname{sen} \nu x),$$

o, lo que es lo mismo,

$$S_n(x) = \sum_{\nu=-n}^{\nu=n} \alpha_\nu e^{i\nu}:$$

Nuestra tarea consiste en demostrar que estas sumas de Fourier son realmente convergentes para $n \to \infty$ y que el límite es la función $f(x)$.

Enunciamos ahora el

TEOREMA PRINCIPAL. *La serie de Fourier*

$$\tfrac{1}{2}a_0 + \sum_{\nu=1}^{\infty} (a_\nu \cos \nu x + b_\nu \operatorname{sen} \nu x), \tag{27a}$$

o sea,

$$\sum_{\nu=-\infty}^{\infty} \alpha_\nu e^{i\nu x}, \tag{27b}$$

formada con los coeficientes de Fourier (26a) *o* (26b), *converge al valor* $f(x)$ *para cualquier función continua por trozos,* $f(x)$, *de período* 2π *y con derivadas de primero y segundo órdenes continuas por trozos.*[1] *Aquí el valor de* $f(x)$ *en un punto de discontinuidad debe definirse como*

$$f(x) = \tfrac{1}{2}[f(x+0) + f(x-0)]. \tag{27c}$$

DEMOSTRACIÓN.[2] Para la demostración, introducimos en el n-ésimo "polinomio de Fourier",

$$S_n(x) = \tfrac{1}{2}a_0 + \sum_{\nu=1}^{\infty} (a_\nu \cos \nu x + b_\nu \operatorname{sen} \nu x),$$

[1] Nuevamente indicamos que este teorema puede demostrarse para clases de funciones mucho más generales (ver, por ejemplo, la sección 8.6). Sin embargo, el resultado que aquí se formula es suficiente para la mayoría de las aplicaciones.

[2] Damos aquí únicamente la demostración para el desarrollo de f en la serie (27a). La serie (27b) se deduce entonces por la substitución dada en la Ec. (13b), página 601.

las expresiones integrales (26a) de los coeficientes e intercambiamos la integración y la suma; y así resulta

$$S_n(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\left[\tfrac{1}{2} + \sum_{\nu=1}^{\infty} (\cos \nu t \cos \nu x + \operatorname{sen} \nu t \operatorname{sen} \nu x)\right] dt,$$

o, usando el teorema de adición para el coseno,

$$S_n(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\left[\tfrac{1}{2} + \sum_{\nu=1}^{n} \cos \nu (t - x)\right] dt.$$

Por la fórmula de suma (14) de la p. 603, se llega a

$$(28) \qquad S_n(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\,\frac{\operatorname{sen}[(n + \frac{1}{2})(t - x)]}{\operatorname{sen}\frac{1}{2}(t - x)}\,dt.$$

Finalmente, haciendo $\tau = t - x$ y recordando que la periodicidad permite desplazar el intervalo de integración en la cantidad x (ver p. 590), resulta

$$(28a) \qquad S_n(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x + \tau)\,\frac{\operatorname{sen}(n + \frac{1}{2})\tau}{\operatorname{sen}\frac{1}{2}\tau}\,d\tau,$$

donde, por supuesto, x tiene un valor fijo.

Ahora demostraremos que $S_n(x)$ tiende a $f(x)$ cuando $n \to \infty$. En efecto,

$$(29) \qquad \lim_{n\to\infty} S_n(x) = \lim_{n\to\infty} \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x + t)\,\frac{\operatorname{sen}(n + \frac{1}{2})t}{\operatorname{sen}\frac{1}{2}t}\,dt = f(x).$$

Como $f(x) = \frac{1}{2}[f(x + 0) + f(x - 0)]$ para todo x, se tiene [ver fórmula (15), p. 602]

$$S_n(x) - f(x) = \frac{1}{\pi}\int_{0}^{\pi} \frac{[f(x + t) - f(x + 0)]}{2 \operatorname{sen}\frac{1}{2}t}\operatorname{sen}(n + \tfrac{1}{2})t\,dt$$
$$+ \frac{1}{\pi}\int_{-\pi}^{0} \frac{[f(x + t) - f(x - 0)]}{2 \operatorname{sen}\frac{1}{2}t}\operatorname{sen}(n + \tfrac{1}{2})t\,dt.$$

Si podemos demostrar que las funciones $[f(x + t) - f(x + 0)]/(2 \operatorname{sen}\frac{1}{2}t)$ y $[f(x + t) - f(x - 0)]/(2 \operatorname{sen}\frac{1}{2}t)$ de la variable t, junto con sus primeras derivadas, son continuas a trozos en los intervalos $0 \leq t \leq \pi$ y $-\pi \leq t \leq 0$ respectivamente, entonces, por el Lema básico (p. 603), las dos integrales que aparecen en el lado derecho de la ecuación anterior tienden a cero cuando $n \to \infty$, quedando así establecida la validez de la fórmula (29).

Por consiguiente, el teorema principal quedará demostrado si podemos demostrar que para un valor fijo de x la función de t definida por

$$\phi(t) = \frac{f(x + t) - f(x + 0)}{2 \operatorname{sen}\frac{1}{2}t} \qquad \text{para } 0 < t < \pi$$

$$\phi(t) = \frac{f(x + t) - f(x - 0)}{2 \operatorname{sen}\frac{1}{2}t} \qquad \text{para } -\pi < t < 0$$

es continua a trozos y tiene primera y segunda derivadas que también lo son, siempre que f, f' y f'' sean continuas a trozos.

Para indagar si el cociente $\phi(t)$ satisface estas condiciones, observamos primero que el denominador se anula solamente para $t = 0$ y que, por consiguiente, ϕ y su primera derivada son continuas a trozos excepto, posiblemente, cerca de $t = 0$. Sólo en el punto singular $t = 0$ podría no tenerse derivabilidad. Todo lo que tenemos que hacer, por consiguiente, es demostrar que $\phi(t)$ y su derivada $\phi'(t)$ se aproximan a ciertos límites si t tiende a cero, respectivamente, desde valores positivos y negativos. Demostraremos, en efecto, que estos límites existen y que tienen los valores

$$\phi(+0) = f'(x+0), \qquad \varphi(-0) = f'(x-0)$$

y

$$\phi'(+0) = \tfrac{1}{2}f''(x+0), \qquad \phi'(-0) = \tfrac{1}{2}f''(x-0),$$

respectivamente.

Para la demostración, introduzcamos la función $g(t)$ por medio de la ecuación $\phi(t) = g(t)h(t)$, donde el factor $h(t)$ está definido por

$$h(t) = \frac{t}{2\operatorname{sen}(t/2)} \quad \text{para } t \neq 0, \qquad h(0) = 1.$$

Tenemos (ver capítulo 5, p. 482) en $h(t)$ una función continua con una derivada continua en todo el intervalo $-\pi \leq t \leq \pi$, siendo $h(0) = 1$, y $h'(0) = 0$. Por consiguiente, al tomar límites haciendo $t \to 0$, los valores de $g(t)$ y $\phi(t)$ coinciden entre sí, y lo mismo ocurre con $g'(t)$ y $\phi'(t) = gh' + g'h$.

Ahora bien, en el intervalo $0 < t < \pi$ (ver comentarios generales acerca de expresiones indeterminadas en el capítulo 5, p. 481) el teorema del valor medio del Cálculo asegura que

$$g(t) = \frac{f(x+t) - f(x+0)}{t} = f'(x+\xi),$$

siendo ξ un valor entre 0 y t. Entonces, cuando t (y por tanto, también cuando ξ) tiende a cero, resulta

$$g(+0) = f'(x+0).$$

Y para la derivada,

$$g'(t) = \frac{tf'(x+t) + f(x+0) - f(x+t)}{t^2},$$

expresión en la cual también el numerador y el denominador son funciones que tienden a cero para $t \to 0$, y cuyas derivadas son, respectivamente, $tf''(x+t)$ y $2t$. Para determinar el límite cuando $t \to 0$ hacemos uso del teorema generalizado del valor medio (ver p. 243); y así encontramos

$$g'(t) = \frac{\eta f''(x+\eta)}{2\eta} = \tfrac{1}{2}f''(x+\eta),$$

donde η es un valor entre 0 y t. Si $t \to 0$, entonces $\eta \to 0$, y, por tanto, como se dijo antes, $g'(+0) = \phi'(+0) = \frac{1}{2}f''(x+0)$.

El mismo razonamiento es aplicable a valores negativos de t. En consecuencia, está justificada la aplicación del Lema, y el teorema principal queda así demostrado.

Puede decirse que el resultado obtenido es lo suficientemente amplio como para cubrir todas las necesidades que surgen en el Cálculo y sus aplicaciones. Sin embargo, el interés teórico de los matemáticos, comenzando con el trabajo original de Dirichlet, se orientó frecuentemente hacia una mayor generalidad, es decir, el desarrollo de funciones pertenecientes a una clase más amplia.[1] Estos esfuerzos han estimulado un análisis más refinado de los conceptos de función e integral y conducido al desarrollo del análisis avanzado de Fourier, campo atractivo y especializado que está fuera del objeto de este libro.

8.5 Ejemplos de series de Fourier

a. Comentarios preliminares

En lo que sigue supondremos que el período de las funciones $f(x)$ es 2π.

Si $f(x)$ es una función par (ver p. 53), es claro entonces que $f(x)$ sen νx es impar y $f(x)$ cos νx es par, de manera que

$$b_\nu = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \operatorname{sen} \nu x \, dx = 0,$$

y se obtiene una "serie de cosenos". Si, por otra parte, la función $f(x)$ es impar, entonces

$$a_\nu = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \cos \nu x \, dx = 0,$$

y lo que resulta es una "serie de senos".[2]

---

[1] Debe hacerse notar que existen ejemplos de funciones continuas que no son desarrollables en serie de Fourier. Además, existen también ejemplos de funciones $f(x)$ representadas por series trigonométricas convergentes que, no obstante, no son series de Fourier con coeficientes dados por las expresiones (26). Tales ejemplos muestran que debe establecerse una distinción entre las series trigonométricas, en general, y las series de Fourier, en particular, sobre todo cuando se trata de investigaciones refinadas. Además, tales ejemplos ilustran el hecho de que, sin duda, condiciones aún más restrictivas que la continuidad son convenientes; no obstante que las restricciones supuestas en nuestro teorema general y en una extensión que se da en la sección 8.6 son bastante más fuertes que las realmente necesarias. (Para la teoría general véase: *Trigonometrical Series,* por A. Zygmund, Chelsea Publishing Co., 1952.)

[2] En consecuencia, si la función $f(x)$ está dada inicialmente sólo en el intervalo $0 < x < \pi$, podemos extenderla al intervalo $-\pi < x < 0$, ya sea como una función impar o como una función par. Así, para el intervalo $0 < x < \pi$, más pequeño, resulta una serie de senos o una serie de cosenos.

b. Desarrollo de la función $\phi(x) = x^2$

Para la función par x^2 se obtiene, después de una doble integración por partes,

$$a_\nu = \frac{2}{\pi}\int_0^\pi x^2 \cos \nu x \, dx = (-1)^\nu \frac{4}{\nu^2}, \qquad (\nu > 0),$$

$$a_0 = \frac{2\pi^2}{3},$$

de manera que resulta el desarrollo

$$x^2 = \frac{\pi^2}{3} - 4\left(\frac{\cos x}{1^2} - \frac{\cos 2x}{2^2} + \frac{\cos 3x}{3^2} + \cdots \right). \tag{30}$$

Derivando término a término esta serie y dividiendo por 2 el resultado, se obtiene formalmente la serie (23*b*) de p. 608, que antes obtuvimos para $\phi(x) = x$.

c. Desarrollo de $x \cos x$

(Véase Fig. 8.9.) Para esta función impar tenemos que

$$a_\nu = 0, \qquad b_\nu = \frac{2}{\pi}\int_0^\pi x \cos x \operatorname{sen} \nu x \, dx.$$

Usando la fórmula

$$\int_0^\pi x \operatorname{sen} \mu x \, dx = (-1)^{\mu+1} \frac{\pi}{\mu}, \qquad (\mu = 1, 2, \ldots,),$$

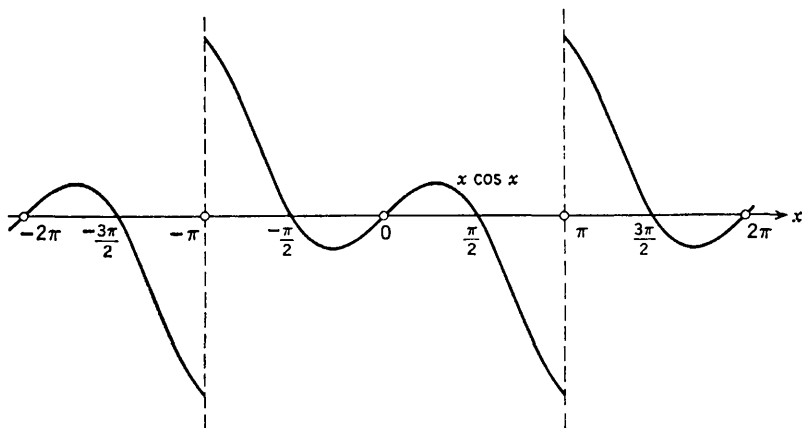


Figura 8.9

encontramos que

$$\begin{aligned} b_\nu &= \frac{2}{\pi}\int_0^\pi x\cos x \operatorname{sen} \nu x\, dx \\ &= \frac{1}{\pi}\int_0^\pi x[\operatorname{sen}(\nu+1)x + \operatorname{sen}(\nu-1)x]\, dx \\ &= (-1)^\nu\left(\frac{2\nu}{\nu^2-1}\right) \qquad (\nu = 2, 3, \ldots) \\ b_1 &= -\tfrac{1}{2}. \end{aligned}$$

Por consiguiente, la serie buscada es

$$(31) \qquad x\cos x = -\tfrac{1}{2}\operatorname{sen} x + 2\sum_{\nu=2}^{\infty}\frac{(-1)^\nu \nu}{\nu^2-1}\operatorname{sen}\nu x.$$

Si a esta serie le sumamos la (23*b*) de p. 608, que obtuvimos para $\phi(x) = x$, obtenemos el resultado

(31*a*)

$$x(1+\cos x) = \tfrac{3}{2}\operatorname{sen} x + 2\left(\frac{\operatorname{sen} 2x}{1\cdot 2\cdot 3} - \frac{\operatorname{sen} 3x}{2\cdot 3\cdot 4} + \frac{\operatorname{sen} 4x}{3\cdot 4\cdot 5} - + \cdots\right).$$

Cuando la función que coincide con $x\cos x$ en el intervalo $-\pi < x < \pi$ se extiende periódicamente más allá de este intervalo, aparecen las mismas discontinuidades (ver Fig. 8.7) que las que exhibe la función $\phi(x)$ considerada antes en la sección 8.4*d*. Por otra parte, la función $x(1+\cos x)$, extendida periódicamente, sigue siendo continua en los puntos extremos del intervalo y, de hecho, su derivada también sigue siendo continua, pues las discontinuidades son eliminadas por el factor $1 + \cos x$, el cual, lo mismo que su derivada, se anula en los puntos extremos. Esto equivale al hecho de que la serie (31) converge uniformemente para todo x, lo cual es evidente si se compara con la serie de términos constantes $\frac{1}{1^3} + \frac{1}{2^3} + \frac{1}{3^3} + \cdots$.

d. La función $f(x) = |x|$

Para esta función par se tiene $b_\nu = 0$ y $a_\nu = \frac{2}{\pi}\int_0^\pi x\cos\nu x\, dx$. Integrando por partes se encuentra de inmediato que

$$\int_0^\pi x\cos\nu x\, dx = \frac{1}{\nu}x\operatorname{sen}\nu x\Big|_0^\pi - \frac{1}{\nu}\int_0^\pi \operatorname{sen}\nu x\, dx$$

$$= \begin{cases} 0, & \text{si } \nu \text{ es par y} \neq 0, \\ -\dfrac{2}{\nu^2}, & \text{si } \nu \text{ es impar.} \end{cases}$$

En consecuencia,

$$|x| = \tfrac{1}{2}\pi - \frac{4}{\pi}\left(\cos x + \frac{\cos 3x}{3^2} + \frac{\cos 5x}{5^2} + \cdots\right). \tag{32}$$

Haciendo $x = 0$ resulta la notable fórmula

$$\frac{\pi^2}{8} = 1 + \frac{1}{3^2} + \frac{1}{5^2} + \cdots. \tag{32a}$$

e. Una función constante por trozos

La función definida por las ecuaciones

$$f(x) = \operatorname{sgn} x = \begin{cases} -1, & \text{para } -\pi < x < 0, \\ 0, & \text{para } x = 0, \\ +1, & \text{para } 0 < x < \pi, \end{cases}$$

como se indica en la Fig. 1.22, p. 56, es impar. Por consiguiente $a_\nu = 0$ y

$$b_\nu = \frac{2}{\pi}\int_0^\pi \operatorname{sen} \nu x\, dx = \begin{cases} 0 & \text{si } \nu \text{ es par}, \\ \dfrac{4}{\pi\nu}, & \text{si } \nu \text{ es impar}, \end{cases}$$

de manera que la serie de Fourier para esta función es

$$f(x) = \frac{4}{\pi}\left(\frac{\operatorname{sen} x}{1} + \frac{\operatorname{sen} 3x}{3} + \cdots\right). \tag{33}$$

En particular, para $x = \tfrac{1}{2}\pi$ se obtiene nuevamente la serie de Leibnitz.

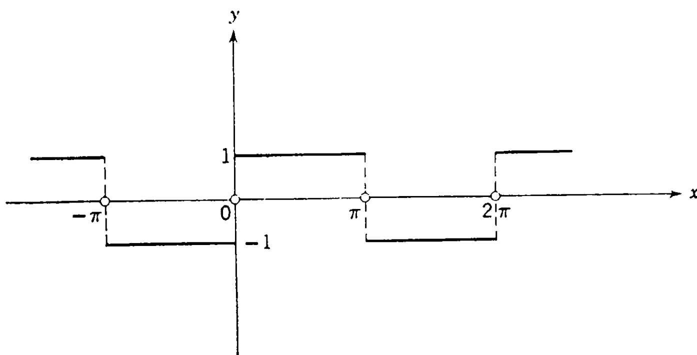


Figura 8.10

La serie (33) puede derivarse formalmente de la serie para $|x|$ dada en (32), por medio de la derivación término a término.

f. La función $|\text{sen}\, x|$

La función par $f(x) = \text{sen}\, x$ puede desarrollarse en una "serie de cosenos", estando dados los coeficientes a_ν por los siguientes cálculos:

$$\begin{aligned}
\tfrac{1}{2}\pi a_\nu &= \int_0^\pi \text{sen}\, x \cos \nu x\, dx \\
&= \tfrac{1}{2}\int_0^\pi [\text{sen}\,(\nu+1)x - \text{sen}\,(\nu-1)x]\, dx \\
&= \begin{cases} 0 & \text{si } \nu \text{ es impar,} \\ \dfrac{-2}{\nu^2-1} & \text{si } \nu \text{ es par.} \end{cases}
\end{aligned}$$

En esta forma resulta, escribiendo 2ν en vez de ν,

$$|\text{sen}\, x| = \frac{2}{\pi} - \frac{4}{\pi}\sum_{\nu=1}^{\infty} \frac{\cos 2\nu x}{4\nu^2 - 1}. \tag{34}$$

g. Desarrollo de $\cos \mu x$. Descomposición de la cotangente en fracciones parciales. El producto infinito para el seno

La función $f(x) = \cos \mu x$ para $-\pi < x < \pi$, en la cual μ no es un entero, es una función par; en consecuencia, $b_\nu = 0$ y

$$\begin{aligned}
\tfrac{1}{2}\pi a_\nu &= \int_0^\pi \cos \mu x \cos \nu x\, dx \\
&= \tfrac{1}{2}\int_0^\pi [\cos\,(\mu+\nu)x + \cos\,(\mu-\nu)x]\, dx \\
&= \tfrac{1}{2}\left[\frac{\text{sen}\,(\mu+\nu)\pi}{u+\nu} + \frac{\text{sen}\,(\mu-\nu)\pi}{\mu-\nu}\right] \\
&= \frac{\mu(-1)^\nu}{\mu^2-\nu^2}\,\text{sen}\,\mu\pi.
\end{aligned}$$

En esta forma tenemos que

$$\cos \mu x = \frac{2\mu\, \text{sen}\, \mu\pi}{\pi}\left(\frac{1}{2\mu^2} - \frac{\cos x}{\mu^2 - 1^2} + \frac{\cos 2x}{\mu^2 - 2^2} + - \cdots\right). \tag{35}$$

Esta función con período 2π, extendida periódicamente fuera del intervalo $-\pi < x < \pi$, sigue siendo continua en los puntos $x = \pm\pi$. Haciendo $x = \pi$, dividiendo ambos lados de la ecuación por $\text{sen}\, \mu\pi$ y escribiendo x en vez de μ, obtenemos la ecuación

$$\cot \pi x = \frac{2x}{\pi}\left(\frac{1}{2x^2} + \frac{1}{x^2 - 1^2} + \frac{1}{x^2 - 2^2} + \cdots\right). \tag{36}$$

Ésta es la descomposición de la cotangente en fracciones parciales (por analogía con la descomposición finita de funciones racionales en fracciones parciales, discutida en el capítulo 3, p. 306). Tal descomposición constituye un resultado muy importante en el Análisis.

Podemos escribir la serie anterior en la forma

$$\cot \pi x - \frac{1}{\pi x} = -\frac{2x}{\pi}\left[\frac{1}{1^2 - x^2} + \frac{1}{2^2 - x^2} + \cdots\right].$$

Si x está en el intervalo $0 \leq x \leq q < 1$, el valor absoluto del n-ésimo término de la derecha es menor que $2/[\pi(n^2 - q^2)]$. Por tanto, la serie converge uniformemente en este intervalo y puede ser integrada término a término. Multiplicando ambos lados por π e integrando resulta en el lado izquierdo

$$\pi \int_0^x \left(\cot \pi t - \frac{1}{\pi t}\right) dt = \log \frac{\operatorname{sen} \pi x}{\pi x} - \lim_{a \to 0} \log \frac{\operatorname{sen} \pi a}{\pi a} = \log \frac{\operatorname{sen} \pi x}{\pi x}$$

y en el derecho

$$\log\left(1 - \frac{x^2}{1^2}\right) + \log\left(1 - \frac{x^2}{2^2}\right) + \cdots = \lim_{n \to \infty} \sum_{\nu=1}^{n} \log\left(1 - \frac{x^2}{\nu^2}\right).$$

Así,

$$\log \frac{\operatorname{sen} \pi x}{\pi x} = \lim_{n \to \infty} \sum_{\nu=1}^{n} \log\left(1 - \frac{x^2}{\nu^2}\right)$$

$$= \lim_{n \to \infty} \log \prod_{\nu=1}^{n}\left(1 - \frac{x^2}{\nu^2}\right) = \log \lim \prod_{\nu=1}^{n}\left(1 - \frac{x^2}{\nu^2}\right).$$

Al pasar del logaritmo a la función exponencial se obtiene

$$\operatorname{sen} \pi x = \pi x\left(1 - \frac{x^2}{1^2}\right)\left(1 - \frac{x^2}{2^2}\right)\left(1 - \frac{x^2}{3^2}\right)\cdots. \tag{36a}$$

Hemos así obtenido la famosa expresión para el seno, consistente en un producto infinito.[1]

De este resultado haciendo $x = \frac{1}{2}$, se obtiene el producto de Wallis

$$\tfrac{1}{2}\pi = \prod_{\nu=1}^{\infty} \frac{2\nu}{2\nu - 1} \cdot \frac{2\nu}{2\nu + 1} = \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdots$$

tal como se obtuvo en la p. 301.

[1] Esta fórmula es particularmente interesante porque exhibe en forma directa la anulación de la función $\operatorname{sen} \pi x$ en los puntos $x = 0, \pm 1, \pm 2, \ldots$. En este aspecto, se asemeja a la factorización de un polinomio cuando se conocen sus raíces.

h. Ejemplos adicionales

Mediante breves cálculos análogos a los anteriores es posible obtener otros ejemplos de desarrollos.

La función $f(x)$ definida por la ecuación $f(x) = \operatorname{sen} \mu x$ para $-\pi < x < \pi$ puede desarrollarse según la serie

$$(37)\quad \operatorname{sen} \mu x = -\frac{2 \operatorname{sen} \mu\pi}{\pi}\left(\frac{\operatorname{sen} x}{\mu^2 - 1^2} - \frac{2 \operatorname{sen} 2x}{\mu^2 - 2^2} + \frac{3 \operatorname{sen} 3x}{\mu^2 - 3^2} - + \cdots\right).$$

Haciendo $x = \frac{1}{2}\pi$ y usando la relación $\operatorname{sen} \mu\pi = 2 \operatorname{sen} \frac{1}{2}\mu\pi \cos \frac{1}{2}\mu\pi$ obtenemos la descomposición de la secante, es decir, de la función $1/\cos \frac{1}{2}\mu\pi$, en fracciones parciales. Tal desarrollo es

$$\pi \sec \pi x = \frac{\pi}{\cos \pi x} = 4 \sum_{\nu=1}^{\infty} \frac{(-1)^{\nu}(2\nu - 1)}{4x^2 - (2\nu - 1)^2},$$

donde en lugar de $\frac{1}{2}\mu$ hemos escrito x.

Las series análogas a la (35) y a la (36) para las funciones hiperbólicas $\cosh \mu x$ y $\operatorname{senh} \mu x (-\pi < x < \pi)$ son

$$\cosh \mu x = \frac{2\mu}{\pi} \operatorname{senh} \mu\pi \left(\frac{1}{2\mu^2} - \frac{\cos x}{\mu^2 + 1^2} + \frac{\cos 2x}{\mu^2 + 2^2} - \frac{\cos 3x}{\mu^2 + 3^2} + \cdots\right),$$

$$\operatorname{senh} \mu x = \frac{2}{\pi} \operatorname{senh} \mu\pi \left(\frac{\operatorname{sen} x}{\mu^2 + 1^2} - \frac{2 \operatorname{sen} 2x}{\mu^2 + 2^2} + \frac{3 \operatorname{sen} 3x}{\mu^2 + 3^2} - + \cdots\right).$$

8.6 Discusión adicional sobre la convergencia

a. Resultados

Un examen más minucioso de los coeficientes de Fourier, a_ν y b_ν, conduce fácilmente a los siguientes corolarios del teorema principal de la sección 8.4*e*, p. 609.

(*a*) La serie de Fourier (27) de p 610 converge a $f(x)$ si $f(x)$ es una función periódica que satisface las condiciones menos estrictas de ser ella misma y su *primera* derivada, $f'(x)$, continuas a trozos o, como también se dice, si $f(x)$ es una función *"lisa" por trozos.*

(*b*) Si la función periódica lisa por trozos, $f(x)$, es continua, la convergencia es absoluta y uniforme.

(*c*) Si la función lisa por trozos, $f(x)$, presenta discontinuidades de salto, la convergencia es uniforme en cualquier intervalo cerrado que no contenga un punto de discontinuidad.

La demostración de (*b*) se basa en una sencilla desigualdad debida a Bessel; en tanto que para demostrar (*a*) y (*c*) se usarán los resultados de la sección 8.4*d*, p. 608.

b. Desigualdad de Bessel

Esta desigualdad proporciona cotas para los coeficientes de Fourier de cualquier función continua a trozos pero no necesariamente derivable. Establece que

$$\tfrac{1}{2}a_0^2 + \sum_{\nu=1}^{n}(a_\nu^2 + b_\nu^2) \leq M^2; \tag{38}$$

donde la cota $M^2 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)^2\,dx$ es un número fijado por la función $f(x)$, que no depende de los coeficientes de Fourier individuales, a_ν, b_ν, ni del número n. Con los coeficientes complejos de Fourier, α_ν [ver (13*a*), p. 601], la desigualdad de Bessel puede escribirse de inmediato en la forma

$$\sum_{\nu=-n}^{n} |\alpha_\nu|^2 \leq \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)^2\,dx = \tfrac{1}{2}M^2. \tag{38a}$$

La desigualdad es una consecuencia directa de otra cuya validez es casi evidente:

$$\frac{1}{\pi}\int_{-\pi}^{\pi}\left[f(x) - \tfrac{1}{2}a_0 - \sum_{\nu=1}^{n}(a_\nu\cos\nu x + b_\nu \operatorname{sen}\nu x)\right]^2 dx \geq 0.$$

La integral aquí se evalúa desarrollando el cuadrado que aparece como integrando y teniendo en cuenta las relaciones de ortogonalidad (25) (p. 609), así como también las definiciones (17) (p. 602) de los coeficientes de Fourier. Al integrar los términos individualmente se llega de inmediato a la desigualdad de Bessel en la forma (38) dada.

Puesto que el lado izquierdo de la desigualdad de Bessel crece monótonamente con n en tanto que la cota superior M^2 está fija, podemos pasar al límite con $n \to \infty$ e inferir la validez de la desigualdad

$$2\sum_{\nu=-\infty}^{\infty}|\alpha_\nu|^2 = \tfrac{1}{2}a_0^2 + \sum_{\nu=1}^{\infty}(a_\nu^2 + b_\nu^2) \leq M^2 \tag{39}$$

Esta última se cumple para los coeficientes de Fourier de una función continua a trozos $f(x)$ aun cuando f pueda no ser representable por las series (27*a*) o (27*b*).

En la sección 8.7*d* demostraremos que la desigualdad de essel (39) sigue siendo válida si se reemplaza el signo de desigualdad por el de igualdad.

*c. Demostración de los corolarios (a), (b) y (c)

Suponiendo que $f(x)$ es continua, aplicamos la desigualdad de Bessel a su derivada, $g(x) = f'(x)$, que es continua a trozos y cuyos coeficientes de Fourier son $c_\nu = +\nu b_\nu$, $d_\nu = -\nu a_\nu$, como puede verificarse fácilmente in-

tegrando por partes (ya que los términos integrados se anulan mutuamente):

$$c_\nu = \frac{1}{\pi}\int_{-\pi}^{\pi} f'(x)\cos \nu x\, dx = +\int_{-\pi}^{\pi} \nu f(x) \operatorname{sen} \nu x\, dx = +\nu b_\nu,$$

y análogamente para d_ν. [Aquí se ha usado la propiedad de que $f(x)$ es continua y periódica.] Se tiene, en consecuencia,

$$\sum_{\nu=1}^{n} \nu^2(a_\nu^2 + b_\nu^2) = \sum_{\nu=1}^{n} (c_\nu^2 + d_\nu^2)$$

$$\leq \frac{1}{\pi}\int_{-\pi}^{\pi} g(x)^2\, dx = \frac{1}{\pi}\int_{-\pi}^{\pi} f'(x)^2\, dx = M^2.$$

Este resultado permite construir para la serie de Fourier de $f(x)$ una serie mayorante con términos positivos constantes, la cual, de acuerdo con la p. 551, asegura la convergencia absoluta y uniforme que se menciona en (*b*). En efecto, para la ν-ésima oscilación armónica la desigualdad de Cauchy-Schwarz (ver p. 39) da

$$|a_\nu \cos \nu x + b_\nu \operatorname{sen} \nu x|^2 \leq (a_\nu^2 + b_\nu^2)(\cos^2 \nu x + \operatorname{sen}^2 \nu x) = a_\nu^2 + b_\nu^2;$$

y usando entonces la desigualdad

$$pq \leq \tfrac{1}{2}(p^2 + q^2)$$

con $p = 1/\nu$, $q = \nu\sqrt{a_\nu^2 + b_\nu^2}$, se obtiene que para todo ν es

$$|a_\nu \cos \nu x + b_\nu \operatorname{sen} \nu x| \leq \frac{1}{\nu}\, \nu\sqrt{a_\nu^2 + b_\nu^2}$$

$$\leq \frac{1}{2}\left[\frac{1}{\nu^2} + \nu^2(a_\nu^2 + b_\nu^2)\right].$$

La suma sobre ν del lado izquierdo de esta última desigualdad converge y, por consiguiente, constituye una mayorante. Por tanto, la serie de Fourier

$$\tfrac{1}{2}a_0 + \sum_{\nu=1}^{\infty} (a_\nu \cos \nu x + b_\nu \operatorname{sen} \nu x)$$

converge uniformemente. Su suma es entonces $s(x)$, que es una función continua de x. Para demostrar que realmente $s(x) = f(x)$ usamos el artificio consistente en considerar la función integrada

$$F(x) = \int_{-\pi}^{x} (f(t) - \tfrac{1}{2}a_0)\, dt.$$

Es claro que $F(x)$ es continua para $-\pi \leq x \leq \pi$; además, F tiene en $x = -\pi$ el mismo valor que en $x = \pi$, pues

$$F(\pi) = \int_{-\pi}^{+\pi} f(t)\, dt - \pi a_0 = 0 = F(-\pi).$$

Por tanto, la extensión periódica de F es continua. Puesto que las dos primeras derivadas de F son continuas a trozos, la función F está representada por su serie de Fourier. Por el mismo argumento que se usó antes, basado en la integración por partes, los coeficientes de Fourier de F son $-(1/\nu)b_\nu$ y $(1/\nu)a_\nu$ para $\nu \neq 0$, de manera que

$$F(x) = \tfrac{1}{2}A_0 + \sum_{\nu=1}^{\infty} \frac{1}{\nu}(-b_\nu \cos \nu x + a_\nu \operatorname{sen} \nu x),$$

siendo A_0 algún coeficiente constante. Ahora sabemos que la serie obtenida formalmente por derivación término a término es uniformemente convergente. En consecuencia, es válido derivar término a término formalmente (ver p. 555), resultando así la relación deseada:

$$F'(x) = f(x) - \tfrac{1}{2}a_0 = \sum_{\nu=1}^{\infty} (a_\nu \cos \nu x + b_\nu \operatorname{sen} \nu x).$$

Con el fin de demostrar los asertos restantes en el caso de una función periódica continua a trozos, f, cuya derivada, f', es también continua a trozos, recordemos que, por los resultados anteriores, dichos asertos son válidos para la función periódica $\chi(x)$ de la sección 8.4*d* y, en consecuencia, también para la función $\chi(x - \xi)$, que presenta un salto de 2π en el punto ξ. Si ahora la función $f(x)$ presenta saltos de $\beta_1, \beta_2, \ldots, \beta_m$ en los puntos $\xi_1, \xi_2, \ldots, \xi_m$, entonces $f^*(x) = f(x) - \frac{1}{2\pi}\sum_{i=1}^{m} \beta_i \chi(x - \xi_i)$ satisface las condiciones de (b) y, por tanto, tiene una serie de Fourier uniformemente convergente; con lo cual se demuestra la validez de (a) y (c) para $f(x)$.

d. Orden de magnitud de los coeficientes de Fourier. Derivación de series de Fourier

Las discusiones anteriores relativas a la convergencia ilustran una propiedad general: Los coeficientes de Fourier a_ν, b_ν convergen más rápidamente a cero, conforme $n \to \infty$, cuando $f(x)$ es más lisa, es decir, cuando es mayor el número de derivadas de la función periódica $f(x)$ que son continuas. En forma equivalente, la convergencia de la serie de Fourier es mejor a medida que la función es más lisa. Enunciado en forma precisa: Si la función periódica $f(x)$ tiene derivadas continuas hasta el orden k y una derivada continua a trozos de orden $k+1$, existe entonces una cota B, que sólo depende de $f(x)$ y de k, tal que

(40) $$|a_\nu|, |b_\nu| < \frac{B}{\nu^{k+1}}.$$

La demostración es casi inmediata (véase lo anterior) si se aplica la integración por partes. Por brevedad escribimos en notación compleja

$$a_\nu - ib_\nu = 2\alpha_\nu$$

e integramos por partes sucesivamente hasta que aparezca como integrando el factor $f^{(k+1)}(x)$. Debido a la periodicidad y continuidad de $f(x)$, $f'(x)$, etc., los términos extremos se anulan mutuamente y resulta

$$2\pi\alpha_\nu = \int_{-\pi}^{\pi} f(x)e^{-i\nu x}\,dx = -\frac{i}{\nu}\int_{-\pi}^{\pi} f'(x)e^{-i\nu x}\,dx$$
$$= \cdots = \left(\frac{-i}{\nu}\right)^{k+1}\int_{-\pi}^{\pi} f^{(k+1)}(x)e^{-i\nu x}\,dx.$$

Por tanto, si $\frac{1}{2}B$ es una cota superior para $|f^{(k+1)}(x)|$, entonces $|\alpha_\nu| \leq \frac{1}{2}B/\nu^{k+1}$, lo cual implica las desigualdades (40).

Otro resultado notable es que para $k > 2$ la serie de Fourier puede derivarse término a término $k - 1$ veces, y en esa forma resulta la serie de Fourier para la función derivada. Esto se demuestra observando que todas estas series derivadas tienen como mayorante la serie convergente $B\sum_{\nu=1}^{\infty}\frac{1}{\nu^2}$, por lo que ellas mismas son absoluta y uniformemente convergentes (ver los criterios del capítulo 7, p. 559).

*8.7 Aproximación mediante polinomios trigonométricos y racionales

a. Comentario general sobre las representaciones de funciones

Desde los primeros años del Cálculo se presentó el difícil problema de establecer la forma en que debería modificarse el concepto de función para dar lugar a la posibilidad de "expresiones explícitas". A menudo las funciones no se dan en forma analítica, sino más bien por construcción geométrica o mecánica, o por la descripción *geométrica* de sus gráficas, las que podrían ser de naturaleza diferente en distintos intervalos.

El descubrimiento de las series de Fourier a principios del siglo XIX constituyó un paso de suma trascendencia en la búsqueda de una respuesta para la vieja pregunta. Revelaba que, en efecto, muchas funciones "arbitrarias", sin duda mucho menos restringidas que las "analíticas", podían expresarse como series de Fourier convergentes. Pero las series de Fourier no cubren todas las funciones continuas: como se dijo antes, sin que se demostrara, es posible definir funciones continuas cuyas series de Fourier, formadas con los coeficientes de Fourier, no convergen.

Es de lo más notable el que, abandonando el principio de las series infinitas, según el cual la aproximación se realiza por la sola adición de términos de orden superior, sea posible construir para cualquier función continua, $f(x)$, polinomios de aproximación trigonométricos o racionales, $P_n(x)$, de grado n, tales que cuando $n \to \infty$ convergen uniformemente a la función dada $f(x)$ en un intervalo cerrado.

b. Teorema de aproximación de Weierstrass

Demostraremos los siguientes teoremas, que están estrechamente relacionados entre sí.

a) Si $f(x)$ es una función continua en un intervalo cerrado, I, contenido dentro del intervalo mayor $-\pi < x < \pi$, entonces f puede aproximarse uniformemente en I por medio de un polinomio trigonométrico de período 2π y de orden n suficientemente grande.

b) Cualquier función $f(x)$ continua en un intervalo cerrado, I, puede aproximarse uniformemente en I por medio de un polinomio en x, $P(x)$. Este enunciado, debido a Weierstrass, puede complementarse (ver p. 555) con el corolario:

**c*) Si $f(x)$ posee una derivada continua en I, entonces los polinomios de aproximación pueden escogerse de manera tal que los polinomios derivados, $P_n'(x)$, aproximen uniformemente a la derivada $f'(x)$.

La demostración de *a*) es bastante sencilla. Primero aproximamos $f(x)$ mediante una función lineal por trozos cuya gráfica es un polígono, $L_n(x)$, inscripto en la gráfica de $f(x)$ (ver Fig. 8.11). Es claro que el valor abso-

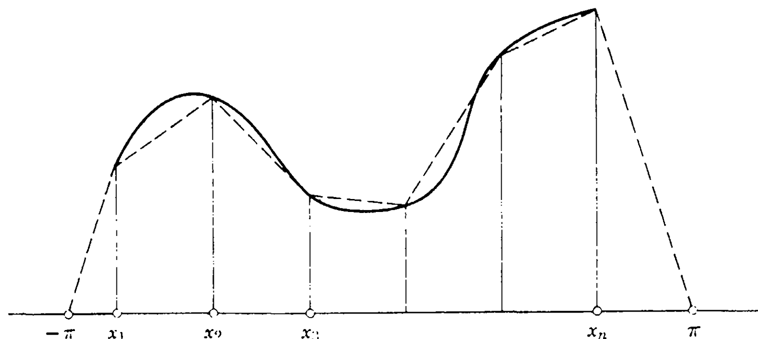


Figura 8.11 Aproximación uniforme de una función continua mediante un polígono.

luto de la diferencia entre $L_n(x)$ y $f(x)$ es menor que cualquier margen dado arbitrariamente pequeño, $\varepsilon/2$, siempre que los vértices del polígono estén sobre los puntos equidistantes $x_1, x_2, \ldots, x_n$ y que la constante $h = x_{\nu+1} - x_\nu$ sea suficientemente pequeña. Esto es una consecuencia de la continuidad uniforme de la función continua f en el intervalo I (ver p. 123).

El siguiente paso es unir, como se indica en la figura, los puntos extremos $-\pi$ y π del intervalo mayor por medio de líneas rectas, extendiendo así $L_n(x)$ a una función lineal por trozos, que nuevamente llamaremos $L_n(x)$, definida en el intervalo cerrado $-\pi \leq x \leq \pi$. Esta función, que se anula

en ambos puntos extremos, puede ahora extenderse periódicamente y, de acuerdo con la sección 8.6*a,* desarrollarse en una serie de Fourier uniformemente convergente, cuya sección polinomial, $S_m(x)$, difiera de $L_n(x)$, en valor absoluto, en una cantidad menor que $\varepsilon/2$ siempre que m sea suficientemente grande. En estas condiciones es $|S_m - f| \leq |S_m - L_n| + |L_n - f| < \varepsilon$, y así *a*) queda demostrado.[1]

Para demostrar (*b*) reemplacemos en cada término de la suma finita $S_m(x)$, según la sección 5.5*b,* p. 471, las funciones trigonométricas $\cos \nu x$ y $\operatorname{sen} \nu x$ por polinomios de Taylor cuyo residuo sea uniformemente pequeño. A continuación, combinando estas últimas aproximaciones construimos un polinomio, $P_N(x)$, para el cual $|P_N(x) - S_m(x)| < \varepsilon/2$, donde N debe escogerse suficientemente grande para asegurar la aproximación dada por $\varepsilon/2$. Combinando aproximaciones, se tiene que en el intervalo más pequeño es $|P_N(x) - f(x)| < \varepsilon$ siempre que m se escoja de manera que

$$|S_m(x) - f(x)| < \varepsilon/2.$$

*c. Aproximación trigonométrica de Fejer para los polinomios de Fourier, usando medias aritméticas

El teorema *a*) de la sección 8.7*b* puede demostrarse de manera muy simple por medio de una construcción directa y más bien explícita del polinomio de aproximación, proporcionada por el notable teorema, debido a L. Fejer, que enunciamos a continuación.

TEOREMA. *Si $S_n(x)$ es el n-ésimo polinomio de Fourier de una función periódica continua, $f(x)$, entonces la media aritmética*

$$F_n(x) = \frac{S_0(x) + \cdots + S_n(x)}{n+1}$$

converge uniformemente a $f(x)$ cuando $n \to \infty$.

El teorema garantiza la convergencia promediando cualquier oscilación perturbadora que pudiera estar presente en la aproximación de Fourier ordinaria.

DEMOSTRACIÓN. La demostración es similar a la del teorema principal sobre desarrollos de Fourier, pero es más simple debido a que el "núcleo" oscilante $\dfrac{\operatorname{sen}(n+\frac{1}{2})x}{2\operatorname{sen}\frac{1}{2}x}$ que allí aparece se reemplaza en este caso por el "núcleo de Fejer" positivo $S_n(t) = \left(\dfrac{\operatorname{sen}\frac{1}{2}(n+1)t}{2\operatorname{sen}\frac{1}{2}t}\right)^2 \cdot \dfrac{2}{n+1}$. Observe-

[1] El mismo resultado se obtiene cuando I es el intervalo total $-\pi \leq x \leq \pi$, si se supone que $f(\pi) = f(-\pi)$. En este caso se escoge, como antes, un polígono de aproximación, $L_n(x)$, bastando tomar $L_n(-\pi) = L_n(\pi) = f(-\pi) = f(\pi)$.

mos primero que la función $\sigma_n(\alpha) = \frac{1}{2} + \cos\alpha + \cdots + \cos n\alpha$ de la p. 601 puede escribirse en la forma

$$\begin{aligned}\sigma_n(\alpha) &= \frac{\operatorname{sen}(n+\frac{1}{2})\alpha}{2\operatorname{sen}\frac{1}{2}\alpha} = \frac{\operatorname{sen}\frac{1}{2}\alpha\operatorname{sen}(n+\frac{1}{2})\alpha}{2\operatorname{sen}^2\frac{1}{2}\alpha} \\ &= \frac{1}{2}\,\frac{\cos n\alpha - \cos(n+1)\alpha}{1-\cos\alpha},\end{aligned}$$

haciendo uso de las fórmulas de adición para el coseno. Así resulta la fórmula

$$\begin{aligned}\frac{\sigma_0(\alpha)+\sigma_1(\alpha)+\cdots+\sigma_n(\alpha)}{n+1} &= \frac{1}{2(n+1)}\,\frac{1-\cos(n+1)\alpha}{1-\cos\alpha} \\ &= \frac{1}{2(n+1)}\left(\frac{\operatorname{sen}[(n+1)\alpha/2]}{\operatorname{sen}(\alpha/2)}\right)^2 \\ &= s_n(\alpha).\end{aligned}$$

Puesto que, por la definición de $\sigma_n(\alpha)$ [ver (14), p. 601], se tiene

$$\frac{1}{\pi}\int_{-\pi}^{\pi}\sigma_k(\alpha)\,d\alpha = 1,$$

debe tenerse también que

$$\frac{1}{\pi}\int_{-\pi}^{\pi}s_n(\alpha)\,d\alpha = 1.$$

Por otra parte [ver (28*a*), p. 611],

$$S_n(x) = \frac{1}{\pi}\int_{-\pi}^{\pi}f(x+t)\sigma_n(t)\,dt$$

y, en consecuencia,

$$\begin{aligned}F_n(x) &= \frac{1}{\pi(n+1)}\int_{-\pi}^{\pi}f(x+t)[\sigma_0(t)+\cdots+\sigma_n(t)]\,dt \\ &= \frac{1}{\pi}\int_{-\pi}^{\pi}f(x+t)s_n(t)\,dt.\end{aligned}$$

Si δ es cualquier número positivo, entonces

$$\begin{aligned}f(x)-F_n(x) &= \frac{1}{\pi}\int_{-\pi}^{\pi}[f(x)-f(x+t)]s_n(t)\,dt \\ &= \frac{1}{\pi}\int_{-\delta}^{\delta}[f(x)-f(x+t)]s_n(t)\,dt \\ &\quad+\frac{1}{\pi}\int_{\delta}^{\pi}[f(x)-f(x+t)]s_n(t)\,dt \\ &\quad+\frac{1}{\pi}\int_{-\pi}^{-\delta}[f(x)-f(x+t)]s_n(t)\,dt.\end{aligned}$$

Para $f(x)$ continua, la continuidad es uniforme y es posible encontrar un δ tal que $|f(x) - f(x+t)| < \frac{1}{3}\varepsilon$ para todo x en $[-\pi, \pi]$ y para $|t| < \delta$. Además, f es acotada: $|f| < M$. Puesto que, por definición,

$$|s_n(t)| \leq \frac{1}{2(n+1)\,\text{sen}^2(\delta/2)}, \quad \text{para} \quad \delta \leq |t| \leq \pi,$$

usando $s_n \geq 0$ se encuentra que

$$|f(x) - F_n(x)| \leq \frac{\varepsilon}{3\pi}\int_{-\delta}^{\delta} |s_n(t)|\,dt + \frac{2\pi}{\pi}\int_{\delta}^{\pi} |s_n(t)|\,dt + \frac{2\pi}{\pi}\int_{-\pi}^{-\delta} |s_n(t)|\,dt$$

$$\leq \frac{\varepsilon}{3\pi}\int_{-\pi}^{\pi} s_n(t)\,dt + \frac{2\pi}{\pi}\cdot\frac{2\pi}{2(n+1)\,\text{sen}^2(\delta/2)}$$

$$= \frac{\varepsilon}{3} + \frac{2\pi}{(n+1)\,\text{sen}^2(\delta/2)};$$

y entonces es claro que

$$|f(x) - F_n(x)| \leq \varepsilon$$

para n suficientemente grande. El teorema queda así demostrado.

*d. Aproximación en media y relación de Parseval

La proximidad de dos funciones, $g(x)$ y $h(x)$, en su intervalo cerrado, I, en el cual son continuas, puede medirse, con miras a la convergencia uniforme, por medio del valor máximo de $|g(x) - h(x)|$. Llamando *norma máxima* al máximo valor absoluto de una función continua, $\phi(x)$ en I, podemos expresar la convergencia uniforme de una sucesión de funciones f_n a una función f cuando $n \to \infty$ diciendo que la norma máxima de la diferencia $f - f_n$, o $f_n - f$, tiende a cero: $||f - f_n|| \to 0$.

Para aproximaciones de Fourier (así como para otras importantes teorías matemáticas que rebasan los objetivos de este volumen) es natural considerar otra medida o "norma" para la desviación entre dos funciones o, lo que es suficiente, para la "distancia" de la función idénticamente cero a una función $\phi(x)$. Es ésta la "media cuadrática" o *"norma media cuadrática"*, $\mu = ||\phi||$, definida por el valor promedio

$$\mu^2 = \frac{1}{l}\int_I \phi(x)^2\,dx = ||\phi||^2$$

donde l es la longitud del intervalo I. Esta medida es menos precisa que la de la amplitud máxima, debido a que su pequeñez no significa necesariamente que la función es pequeña en todas partes.

Como un ejemplo, la amplitud de x^n en el intervalo I: $0 \leq x \leq 1$ tiene el valor $(2n+1)^{-1/2}$, el cual puede hacerse arbitrariamente pequeño con

sólo escoger n suficientemente grande, en tanto que la función x^n es igual a 1 para $x = 1$.

Si la norma cuadrática $\|f_n - f\|$ tiende a cero cuando $n \to \infty$, decimos entonces que f_n tiende a f *"en media cuadrática"*.

Puede ser de utilidad designar como "distancia" a la norma cuadrática, debido a la llamada desigualdad del triángulo, que corresponde a la que es válida para números (ver p. 38). Esta desigualdad, $\|f + g\| \le \|f\| + \|g\|$, para las dos funciones f y g, es fácil de deducir: aplicando la desigualdad $pq \le \frac{1}{2}(p^2 + q^2)$ con $p = \dfrac{f(x)}{\|f\|}$, $q = \dfrac{g(x)}{\|g\|}$, e integrando sobre I se obtiene

$$\frac{1}{l}\int_I f(x)g(x)\, dx \le \|f\| \cdot \|g\|.$$

Por tanto,

$$\begin{aligned} \|f + g\|^2 &= \frac{1}{l}\int_I [f(x) + g(x)]^2\, dx \\ &= \|f\|^2 + \|g\|^2 + \frac{1}{l}\int_I 2f(x)g(x)\, dx \\ &\le (\|f\| + \|g\|)^2, \end{aligned}$$

o sea,

$$\|f + g\| \le \|f\| + \|g\|.$$

Con estos conceptos podemos hacer más clara la desigualdad de Bessel de la sección 8.6*b*. Primero demostraremos que: *La aproximación en media cuadrática más precisa para una función continua a trozos, $f(x)$, mediante un polinomio trigonométrico de orden n,*

$$\begin{aligned} T_n &= \frac{c_0}{2} + \sum_{\nu=1}^{n} c_\nu \cos \nu x + d_\nu \operatorname{sen} \nu x \\ &= \sum_{\nu=-n}^{n} \beta_\nu e^{i\nu x}, \end{aligned}$$

[con $\beta_0 = c_0/2$, $\beta_\nu + \beta_{-\nu} = c_\nu$, $i(\beta_\nu - \beta_{-\nu}) = d_\nu$], *en la cual la elección de los coeficientes c_ν, d_ν es arbitraria, está dada por el polinomio de Fourier*

$$\begin{aligned} S_n &= \frac{a_0}{2} + \sum_{\nu=1}^{n} a_\nu \cos \nu x + b_\nu \operatorname{sen} \nu x \\ &= \sum_{\nu=-n}^{n} \alpha_\nu e^{i\nu x}, \end{aligned}$$

donde a_ν, b_ν y α_ν son, respectivamente, precisamente los coeficientes reales y complejos de Fourier determinados a partir de f por medio de las fórmulas (26) *de p.* 609.

La demostración, escrita por brevedad en notación compleja, se obtiene fácilmente usando las relaciones de ortogonalidad (25) en el intervalo $I = [-\pi, \pi]$ para las funciones $e^{i\nu x}$:

$$\begin{aligned}
\frac{1}{l}\int_I \left(f(x) - \sum_{\nu=-n}^{n} \beta_\nu e^{i\nu x}\right)^2 dx
&= \frac{1}{l}\int_I \left[f(x)^2 - 2\sum_{\nu=-n}^{n} \beta_\nu f(x) e^{i\nu x} + \left(\sum_{\nu=-n}^{n} \beta_\nu e^{i\nu x}\right)^2\right] dx \\
&= \|f\|^2 - 2\sum_{\nu=-n}^{n} \beta_\nu \alpha_{-\nu} + \sum_{\nu=-n}^{n} \beta_\nu \beta_{-\nu} \\
&= \|f\|^2 - \sum_{\nu=-n}^{n} \alpha_\nu \alpha_{-\nu} + \sum_{\nu=-n}^{n} (\alpha_\nu - \beta_\nu)(\alpha_{-\nu} - \beta_{-\nu}) \\
&= \|f\|^2 - \sum_{\nu=-n}^{n} \alpha_\nu \bar{\alpha}_\nu + \sum_{\nu=-n}^{n} (\alpha_\nu - \beta_\nu)(\bar{\alpha}_\nu - \bar{\beta}_\nu) \\
&= \|f\|^2 - \sum_{\nu=-n}^{n} |\alpha_\nu|^2 + \sum_{\nu=-n}^{n} |\alpha_\nu - \beta_\nu|^2;
\end{aligned}$$

Es claro que esta última expresión alcanza su mínimo cuando los β_ν se toman como los coeficientes de Fourier α_ν, es decir, $\alpha_\nu = \beta_\nu$ o, lo que es lo mismo, $c_\nu = a_\nu$, $d_\nu = b_\nu$.

Ahora podemos demostrar el *teorema de Parseval* usando los resultados de aproximación obtenidos antes.

La *desigualdad de Bessel,*

$$\tfrac{1}{2}a_0^2 + \sum_{\nu=1}^{n} (a_\nu^2 + b_\nu^2) \leq \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)^2\, dx$$

cuando $n \to \infty$ *se convierte en la igualdad de Parseval,*

$$\tfrac{1}{2}a_0^2 + \sum_{\nu=1}^{\infty} (a_\nu^2 + b_\nu^2) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)^2\, dx,$$

válida para cualquier función $f(x)$ *de período* 2π *que sea continua para todo* x.

Demostración. Por el teorema de aproximación de Weierstrass para polinomios trigonométricos, podemos escoger una sucesión de polinomios, T_n tal que $[f(x) - T_n(x)] \to 0$ uniformemente en x. Por tanto, también

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} [f(x) - T_n(x)]^2\, dx \to 0 \quad \text{cuando } n \to \infty.$$

Sin embargo, de acuerdo con nuestro último resultado, el polinomio de Fourier

$$S_n(x) = \frac{a_0}{2} + \sum_{\nu=1}^{n} (a_\nu \cos \nu x + b_\nu \operatorname{sen} \nu x)$$

constituye, entre todos los polinomios de Fourier de orden n, la mejor aproximación en media para $f(x)$, de manera que

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}[f(x)-S_n(x)]^2\,dx = \frac{1}{2\pi}\int_{-\pi}^{\pi}[f(x)-T_n(x)]^2\,dx.$$

Por tanto,

$$\lim_{n\to\infty}\frac{1}{2\pi}\int_{-\pi}^{\pi}[f(x)-S_n(x)]^2\,dx = 0.$$

Al cuadrar el integrando como se hizo en la p. 628, se obtiene la relación de Parseval.

Finalmente, hacemos la observación de que la relación de Parseval sigue siendo válida cuando $f(x)$ tiene un cierto número de discontinuidades de salto. La demostración, aunque sencilla, no la daremos aquí.

Apéndice I

*A.I.1 Alargamiento del intervalo del período. Teorema de la integral de Fourier

El intervalo base, $-\pi \leq x \leq \pi$, de las funciones periódicas podría reemplazarse por cualquier intervalo $-B \leq x \leq B$. Mediante la transformación $y = \pi x/B$ este intervalo de longitud $2B$ se transforma en el intervalo $-\pi \leq y \leq \pi$, y una función $f(x)$ con período $2B$ se transforma en una función $g(y) = f(By/\pi) = f(x)$ con período 2π. El teorema principal, escrito en forma compleja [ver fórmula (27a), p. 610], implica que

$$g(y) = \sum_{\nu=-\infty}^{\infty}\frac{1}{2\pi}\int_{-\pi}^{\pi} g(t)e^{-i\nu(t-y)}\,dt;$$

y por la transformación se tiene entonces que

$$f(x) = \frac{1}{2\pi}\sum_{\nu=-\infty}^{\infty}\frac{\pi}{B}\int_{-B}^{B} f(s)e^{-i\nu\pi(s-x)/B}\,ds, \tag{41}$$

donde la variable de integración se ha reemplazado por $s = Bt/\pi$.

La relación (41) es válida para toda función lisa a trozos en $-B \leq x \leq B$.

Hagamos $\pi/B = h$, $\nu\pi/B = \nu h = u_\nu$ y escribamos (41) en la forma

$$f(x) = \frac{1}{2\pi}\sum_{\nu=-\infty}^{\infty} h\int_{-B}^{B} f(s)e^{-iu_\nu(s-x)}\,ds$$

$$= \frac{1}{2\pi}\sum_{\nu=-\infty}^{\infty} he^{iu_\nu x}H_\nu,$$

donde $H_\nu = \int_{-B}^{B} e^{-iu_\nu s} f(s)\, ds$. El paso formal al límite para $B \to \infty$, o $\Delta u = h \to 0$, es obvio y conduce a

$$(42) \qquad f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{iux}\, du \int_{-\infty}^{-\infty} e^{-ius} f(s)\, ds.$$

Esta es la fórmula de la *integral de Fourier,* que será demostrada rigurosamente en el Volumen II para una amplia clase de funciones f. Tal fórmula puede escribirse en una forma simétrica más clara, en términos de un par de relaciones integrales recíprocas entre una función $f(x)$ y su "transformada de Fourier", $F(u)$:

$$(43) \qquad F(u) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(s) e^{-ius}\, ds$$

$$(43a) \qquad f(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} F(u) e^{iux}\, dx.$$

La fórmula de la integral de Fourier, (42), puede escribirse en una forma que no requiere el uso de exponentes imaginarios. Para ello sólo se necesita hacer uso de las expresiones

$$e^{iux}e^{-ius} = e^{iu(x-s)} = \cos u(s-x) - i \operatorname{sen} u(s-x).$$

Puesto que $\operatorname{sen} u(s-x)$ es una función impar de u y $\cos u(s-x)$ una función par, la integración con respecto a u, desde $-\infty$ hasta $+\infty$, del término con seno da una contribución nula, en tanto que al integrar el término con coseno se obtiene el doble del valor obtenido al integrar desde 0 hasta ∞. Por tanto,

$$(43b) \qquad f(x) = \frac{1}{\pi}\int_{0}^{\infty} du \int_{-\infty}^{+\infty} f(s) \cos u(s-x)\, ds.$$

*A.I.2 Fenómeno de Gibbs en puntos de discontinuidad

La convergencia de la serie de Fourier en la vecindad de una discontinuidad de salto exhibe una característica notable que Gibbs descubrió examinando las gráficas de los polinomios de Fourier

$$S_n(x) = \frac{a_0}{2} + \sum_{\nu=1}^{n} (a_\nu \cos \nu x + b_\nu \operatorname{sen} \nu x).$$

Como se hizo resaltar en el capítulo 7, p. 547, la convergencia no uniforme de una sucesión convergente en la vecindad de una discontinuidad de la función límite puede ser visualizada por la forma en que las gráficas continuas de las funciones de aproximación dejan de ser buenas aproximaciones para la gráfica discontinua de la función límite.

En un desarrollo de Fourier estas gráficas no aproximan en forma simple a la gráfica de $f(x)$ aumentada por los segmentos verticales $x = \xi$ que unen los dos puntos extremos en el sitio, ξ, del salto. En vez de ello, las gráficas de S_n presentan ondas que en la proximidad de ξ exceden a las ordenadas $f(\xi + 0)$ y $f(\xi - 0)$, a uno y otro lado, en casi un 9% de la altura total del salto. Así, las gráficas de aproximación se aproximan a la gráfica de $f(x)$ aumentada por un segmento de recta vertical en $x = \xi$, no sólo uniendo los dos puntos extremos de la gráfica de $f(x)$ sino rebasando el segmento de unión en ambos extremos (ver Figs. 8.4 y 8.5, pp. 595 y 596).

El análisis matemático de esta situación es sencillo y bastará discutirlo solamente para la discontinuidad de la función $\chi(x)$ de la sección 8.6*d*, caso al que se redujeron (p. 624) todas las discontinuidades de salto.

La función $\frac{1}{2}\chi(x)$ para x positivo [ver fórmula (23*a*), p. 607] está dada por

$$\tfrac{1}{2}\chi(x) = \tfrac{1}{2}(\pi - x) = \sum_{\nu=1}^{\infty} \frac{\operatorname{sen} \nu x}{\nu}, \qquad 0 < x < \pi.$$

Integrando la fórmula (14) de p. 601, se encuentra que

$$S_n(x) = \sum_{\nu=1}^{n} \frac{\operatorname{sen} \nu x}{\nu} = -\tfrac{1}{2}x + \int_0^x \frac{\operatorname{sen}(n + \frac{1}{2})t}{2 \operatorname{sen} \frac{1}{2}t}\, dt$$

Por tanto, el residuo $r_n(x) = \frac{1}{2}\chi(x) - S_n(x)$ toma la forma

$$r_n(x) = \tfrac{1}{2}\pi - \int_0^x \frac{\operatorname{sen}(n + \frac{1}{2})t}{t}\, dt + \rho_n(x),$$

donde

$$\rho_n(x) = \int_0^x \frac{2 \operatorname{sen} \frac{1}{2}t - t}{2t \operatorname{sen} \frac{1}{2}t} \operatorname{sen}(n + \tfrac{1}{2})t\, dt.$$

Puesto que la expresión $(2 \operatorname{sen} \frac{1}{2}t - t)/2t \operatorname{sen} \frac{1}{2}t$ es continua a trozos y tiene una primera derivada que también lo es, del lema de la p. 604 inferimos que $\rho_n(x)$ tiende a cero uniformemente cuando $n \to \infty$ para $0 < x < \pi$. Además,

$$\sigma_n(x) = \tfrac{1}{2}\pi - \int_0^x \frac{\operatorname{sen}(n + \frac{1}{2})t}{t}\, dt = \tfrac{1}{2}\pi - \int_0^{(n+\frac{1}{2})x} \frac{\operatorname{sen} t}{t}\, dt$$

tiende a cero cuando $n \to \infty$ para cualquier valor de x positivo individualmente considerado (ver p. 605). La convergencia, sin embargo, no es uniforme. Es claro que la derivada de $\sigma_n(x)$ se anula en los puntos $x_k = 2k\pi/(2n + 1)$, para $k = 1, 2, 3, \ldots$. En forma más precisa, es fácil ver que $\sigma_n(x)$ tiene mínimos en los puntos $x_1, x_3, x_5, \ldots$ y máximos en $x_2, x_4, \ldots$. Más aún, los valores de σ_n en los puntos de mínimos forman una sucesión creciente. Así, $\sigma_n(x)$ tiene como mínimo "absoluto", para x positivos, el valor

$$\begin{aligned}\sigma_n(x_1) &= \tfrac{1}{2}\pi - \int_0^\pi \frac{\operatorname{sen} t}{t}\, dt \\ &= \tfrac{1}{2}\pi - \int_0^\pi \left(1 - \frac{1}{3!}t^2 + \frac{1}{5!}t^4 - \cdots\right) dt \\ &= \pi\left(\tfrac{1}{2} - 1 + \frac{\pi^2}{2\cdot 3\cdot 3} - \frac{\pi^4}{2\cdot 3\cdot 4\cdot 5\cdot 5}\right. \\ &\qquad\left. + \frac{\pi^6}{2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7\cdot 7} - \cdots\right) \\ &\approx -0.090\cdots\pi.\end{aligned}$$

Cuando n es grande el residuo r_n es aproximadamente igual a σ_n. En consecuencia, para n grande el polinomio de aproximación S_n excede a la función χ aproximadamente en $(9/100)\pi$, esto es, en cerca de un 9% de la diferencia entre los valores límite derecho e izquierdo de la función en el origen. Así, la rama oscilante de la gráfica de $S_n(x)$ sobrepasa efectivamente la altura de la gráfica de $\chi(x)$ y exhibe el fenómeno de límite descripto anteriormente.

Es fácil ver que los valores medios de Fejer para la suma $S_n(x)$ están libres del fenómeno de Gibbs.

*A.I.3 Integración de series de Fourier

En general, como hemos visto (p. 552), una serie infinita puede integrarse término a término si es uniformemente convergente. Pero en el caso de las series de Fourier tenemos el notable resultado de que la integración por términos siempre es posible. Afirmamos entonces lo siguiente: *Si $f(x)$ es una función continua a trozos en $-\pi \leq x \leq \pi$, cuyo desarrollo formal de Fourier es*

$$\tfrac{1}{2}a_0 + \sum_{\nu=1}^{\infty} (a_\nu \cos \nu x + b_\nu \operatorname{sen} \nu x),$$

entonces, para dos puntos x_1 y x_2 cualesquiera,

$$\int_{x_1}^{x_2} f(x)\, dx = \int_{x_1}^{x_2} \tfrac{1}{2}a_0\, dx + \sum_{\nu=1}^{\infty}\int_{x_1}^{x_2} (a_\nu \cos \nu x + b_\nu \operatorname{sen} \nu x)\, dx,$$

es decir, la serie de Fourier puede integrarse término a término. Además, la serie del lado derecho converge uniformemente con respecto a x_2 para x_1 fijo.

La parte notable de este teorema es que no sólo no se necesita garantizar la convergencia uniforme de la serie sino que tampoco se necesita hacer uso de su convergencia.

Para demostrar el teorema defínase, como en la p. 623,

$$F(x) = \int_{-\pi}^{x} [f(t) - \tfrac{1}{2}a_0]\,dt.$$

La función $F(x)$ es continua y tiene derivada continua a trozos; además, satisface la condición $F(\pi) = F(-\pi) = 0$, de manera que sigue siendo continua cuando se extiende periódicamente. Así, la serie de Fourier para $F(x)$,

$$\tfrac{1}{2}A_0 + \sum_{\nu=1}^{\infty} (A_\nu \cos \nu x + B_\nu \operatorname{sen} \nu x),$$

converge uniformemente a $F(x)$. Integrando por partes se obtienen, para $\nu \neq 0$, los valores

$$A_\nu = \frac{1}{\pi}\int_{-\pi}^{\pi} F(t) \cos \nu t\,dt = -\frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\frac{\operatorname{sen} \nu t}{\nu}\,dt = -\frac{b_\nu}{\nu};$$

$$B_\nu = \frac{1}{\pi}\int_{-\pi}^{\pi} F(t) \operatorname{sen} \nu t\,dt = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\frac{\cos \nu t}{\nu}\,dt = \frac{a_\nu}{\nu},$$

para los coeficientes de Fourier. Por tanto, la serie

$$F(x_2) - F(x_1) = \sum_{\nu=1}^{\infty} [A_\nu(\cos \nu x_2 - \cos \nu x_1) + B_\nu(\operatorname{sen} \nu x_2 - \operatorname{sen} \nu x_1)]$$

$$= \sum_{\nu=1}^{\infty}\left[-\frac{b_\nu}{\nu}(\cos \nu x_2 - \cos \nu x_1) + \frac{a_\nu}{\nu}(\operatorname{sen} x\nu_2 - \operatorname{sen} \nu x_1)\right]$$

converge uniformemente en x. Reemplazando $F(x)$ por $\int_{\pi}^{x}[f(x) - \tfrac{1}{2}a_0]\,dx$, obtenemos la relación

$$\int_{x_1}^{x_2} [f(x) - \tfrac{1}{2}a_0]\,dx = \sum_{\nu=1}^{\infty}\int_{x_1}^{x_2} (a_\nu \cos \nu x + b_\nu \operatorname{sen} \nu x)\,dx,$$

como se afirmó.

Apéndice II

*A.II.1 Polinomios de Bernoulli y sus aplicaciones

a. Definición y desarrollo de Fourier

En la obtención de la serie de Taylor (p. 467) desempeñaron un papel importante los polinomios en x de grado $n \geq 1$: $P_n(x) = (x - \xi)^n/n!$, con parámetro ξ. La sucesión de estos polinomios está caracterizada por las condiciones de que todo polinomio P_{n+1} es una función primitiva de P_n, es decir, $P'_{n+1}(x) = P_n(x)$, y además que $P_n(\xi) = 0$ y $P_0(x) = 1$.

Construiremos ahora otra notable sucesión de polinomios mediante integraciones sucesivas: los *polinomios de Bernoulli,* que luego extenderemos como funciones periódicas y desarrollaremos en serie de Fourier.

Los polinomios de Bernoulli, $\phi_n(x)$, para $0 \leq x \leq 1$, están definidos por las siguientes relaciones de recurrencia:

$$\phi_n'(x) = \phi_{n-1}(x), \qquad \phi_0(x) = 1 \tag{44a}$$

$$\int_0^1 \phi_n(x)\,dx = 0, \qquad \text{para } n > 0. \tag{44b}$$

Cuando $\phi_0, \phi_1, \ldots, \phi_n$ son conocidas, la condición (44*a*) determina ϕ_n a menos de una constante de integración arbitraria. Esta constante se calcula después usando la condición (44*b*). Por inducción vemos de inmediato que ϕ_n es un polinomio de grado n cuyos coeficientes son números racionales. Los primeros polinomios de Bernoulli son fáciles de calcular:

$$\begin{aligned}
\phi_0(x) &= 1,\\
\phi_1(x) &= x - \frac{1}{2},\\
\phi_2(x) &= \frac{1}{2}x^2 - \frac{1}{2}x + \frac{1}{12},\\
\phi_3(x) &= \frac{1}{6}x^3 - \frac{1}{4}x^2 + \frac{1}{12}x,\\
\phi_4(x) &= \frac{1}{24}x^4 - \frac{1}{12}x^3 + \frac{1}{24}x^2 - \frac{1}{720}.
\end{aligned}$$

Para $n > 1$ las relaciones (44*a, b*) dan

$$\phi_n(1) - \phi_n(0) = \int_0^1 \phi_n'(t)\,dt = 0.$$

Por consiguiente, los polinomios ϕ_n pueden ser extendidos, partiendo del intervalo básico $0 \leq x \leq 1$, a todos los valores de x, originándose así funciones *periódicas* $\psi_n(x)$ con período igual a 1. Éstas son las llamadas *funciones de Bernoulli.* La función $\psi_1(x)$ coincide con la función discontinua $\frac{1}{2\pi}\phi(2\pi x - \pi)$ y puede representarse [ver fórmula (23*b*), p. 608] como una serie de Fourier:

$$\psi_1(x) = \frac{1}{\pi}\left(\frac{\operatorname{sen} 2\pi x}{1} + \frac{\operatorname{sen} 4\pi x}{2} + \frac{\operatorname{sen} 6\pi x}{3} + \cdots\right). \tag{45a}$$

Mediante integraciones sucesivas se obtiene entonces

$$\psi_n(t) = (-1)^{(n/2)+1} \cdot \frac{2}{(2\pi)^n}\sum_{k=1}^{\infty}\frac{\cos 2\pi kt}{k^n}, \qquad \text{para } n \text{ par} \tag{45b}$$

$$\text{(45c)} \qquad \psi_n(t) = (-1)^{(n+1)/2} \cdot \frac{2}{(2\pi)^n} \sum_{k=1}^{\infty} \frac{\operatorname{sen} 2\pi k t}{k^n}, \qquad \text{para } n \text{ impar.}$$

En el intervalo original, $0 \leq x \leq 1$, las funciones periódicas $\psi_n(t)$ son idénticas a los polinomios de Bernoulli $\phi_n(t)$.

Para n, par, ψ_n es una función par; para n impar, ψ_n es impar. En forma equivalente

$$\text{(45d)} \qquad \psi_n(-x) = (-1)^n \psi_n(x).$$

Los términos constantes en los polinomios de Bernoulli sucesivos forman una notable sucesión de números racionales

$$\text{(46a)} \qquad b_n = \phi_n(0).$$

$$= \begin{cases} \psi_n(0) & \text{para } n \neq 1, \\ -\frac{1}{2} & \text{para } n = 1. \end{cases}$$

Del desarrollo de Fourier obtenemos inmediatamente

$$\text{(46b)} \qquad b_n = 0 \qquad \text{para } n = 3, 5, \ldots,$$

$$\text{(46c)} \qquad b_n = (-1)^{(n/2)+1} \cdot \frac{2}{(2\pi)^n} \sum_{k=1}^{\infty} \frac{1}{k^n}, \qquad \text{para } n = 2, 4, \ldots.$$

Además, es evidente que para n par $(n = 2m)$ los signos de b_{2m} se alternan.

En lugar de los números b_n, que decrecen rápidamente al crecer n, Jacobo Bernoulli introdujo los números siguientes, que resultan más adecuados:

$$\text{(47)} \qquad B_m = (-1)^{m-1}(2m)!\, b_{2m}.$$

Éstos son los llamados *números de Bernoulli.* (La igualdad entre los números $B_{2m}{}^* = (-1)^{m-1}B_m$ y los números de Bernoulli introducidos en la p. 577 se hará evidente después.) En particular,

$$B_1 = \frac{1}{6}, \qquad B_2 = \frac{1}{30}, \qquad B_3 = \frac{1}{42}, \qquad B_4 = \frac{1}{30},$$

$$B_5 = \frac{5}{66}, \qquad B_6 = \frac{691}{2730}, \qquad B_7 = \frac{7}{6}, \ldots.$$

Como una consecuencia de la fórmula (46*c*) se tiene

$$\text{(48)} \qquad \sum_{k=1}^{\infty} \frac{1}{k^{2n}} = (-1)^{n-1}(2\pi)^{2n}\tfrac{1}{2}b_{2n} = \frac{(2\pi)^{2n}}{2(2n)!} B_n,$$

la cual es una representación explícita de la función ζ de Riemann, $\zeta(s)$, para enteros $s = 2n$ (ver p. 575), en términos de números conocidos. Con ella es posible obtener, por ejemplo, fórmulas tan notables como

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots = \frac{\pi^2}{6} = \zeta(2)$$

y

$$1 + \frac{1}{2^4} + \frac{1}{3^4} + \frac{1}{4^4} + \cdots = \frac{\pi^4}{90} = \zeta(4).$$

Cuando $n \to \infty$, los números b_n y B_n tienden, respectivamente, a cero e infinito. En efecto, vemos en seguida que

$$1 < \sum_{k=1}^{\infty} \frac{1}{k^{2n}} \leq \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6} < 2.$$

Y, por consiguiente,

$$2(2\pi)^{-2n} < |b_{2n}| < 4(2\pi)^{-2n}.$$

Puesto que $2\pi > 1$ y $(2\pi)^{-2n} \to 0$ cuando $n \to \infty$, resulta que $b_{2n} \to 0$, en tanto que $b_{2n+1} = 0$. Además

$$B_n = (2\pi)!\,|b_{2n}| > 2(2\pi)!\,(2\pi)^{-2n};$$

y, como es fácil ver, el lado derecho tiende a infinito al crecer n.

*b. La función generadora. La serie de Taylor para la cotangente trigonométrica e hiperbólica

Los números y los polinomios de Bernoulli conducen de manera elegante al desarrollo de Taylor para la cotangente y funciones relacionadas. Estos desarrollos se deducen de la manera más fácil por medio de la llamada *función generadora* de las funciones de Bernoulli, a saber, la función

$$F(t, z) = \sum_{n=0}^{\infty} \psi_n(t) z^n. \tag{49}$$

Es ésta una serie de potencias en z cuyos coeficientes son las funciones de Bernoulli que dependen del parámetro t. Sobre la base del desarrollo de Fourier, de las Ecs. (45) se llega a la estimación

$$|\psi_n(t)| \leq \left[\frac{2}{(2\pi)^n}\right] \sum_{k=1}^{\infty} \frac{1}{k^n} \leq \left[\frac{2}{(2\pi)^n}\right] \sum_{k=1}^{\infty} \frac{1}{k^2}$$

$$= \frac{\pi^2}{3(2\pi)^n} < \frac{4}{(2\pi)^n}$$

para todo t y todo $n \geq 2$. Por tanto, el valor absoluto del n-ésimo término de la serie para $F(t, z)$ es menor que $4(|z|/2\pi)^n$. Así, para todo t el radio de convergencia de la serie de potencias en z es por lo menos igual a 2π, como se ve comparando con la serie

$$4 \sum_{m=1}^{\infty} \left(\frac{|z|}{2\pi}\right)^n.$$

Puesto que para un valor fijo de z tal que $|z| < 2\pi$ la serie para $F(z,t)$ tiene una serie mayorante convergente, independiente de t, de la teoría general (ver p. 551) se deduce que la serie converge uniformemente para todo t. Puede entonces ser integrada término a término en este dominio, y su derivación término a término será posible siempre que la serie resultante sea también uniformemente convergente. Usaremos esta propiedad para encontrar una forma explícita que dé $F(t,z)$ (ver p. 555). Derivando término a término con respecto a t obtenemos formalmente para $0 < t < 1$ (si $t = 0$ o $t = 1$, $\psi_1(t)$ no tiene derivada):

$$\begin{aligned}\frac{d}{dt}F(t,z) &= \sum_{n=1}^{\infty} \psi_{n-1}(t) z^n \\ &= z\sum_{n=1}^{\infty} \psi_{n-1}(t) z^{n-1} \\ &= z\sum_{n=0}^{\infty} \psi_n(t) z^n \\ &= zF(t,z).\end{aligned}$$

Esta serie tiene la misma forma que la original y es, en efecto, uniformemente convergente, de manera que la derivación término a término se justifica. En consecuencia, para todo valor fijo de z tal que $|z| < 2\pi$ y para $0 < t < 1$ la función generadora $F(t,z)$ satisface la ecuación diferencial $dF/dt = zF(t,z)$. La solución general de esta ecuación es $F = ce^{zt}$, donde c es un factor cuyo valor depende del parámetro z (ver p. 244). Para determinar c integramos la serie de $F(t,z)$ con respecto a t entre 0 y 1:

$$\begin{aligned}\int_0^1 F(t,z)\, dt &= c\int_0^1 e^{zt}\, dt \\ &= c\,\frac{e^z - 1}{z} \\ &= \int_0^1 \sum_{n=0}^{\infty} z^n \psi_n(t)\, dt \\ &= 1 + \sum_{n=1}^{\infty} z^n \int_0^1 \psi_n(t)\, dt = 1.\end{aligned}$$

Por consiguiente, $c = z/(e^z - 1)$, y así llegamos al resultado final:

$$F(t,z) = \frac{ze^{zt}}{e^z - 1}. \tag{50}$$

Haciendo $t \to 0$ en esta expresión se obtiene la serie de Taylor para la función $z/(e^z - 1)$:

$$\lim_{t\to 0} F(t,z) = \frac{z}{e^z - 1} = 1 + \sum_{n=1}^{\infty} b_n z^n.$$

Puesto que $b_1 = -\frac{1}{2}$, al sumar $\frac{1}{2}z$ a ambos lados de esa ecuación resulta

$$\frac{z}{e^z - 1} + \frac{z}{2} = 1 + \sum_{n=2}^{\infty} b_n z^n. \tag{51}$$

De paso se nota que esta fórmula prueba que los números $B_n^* = n!b_n$ son los números de Bernoulli introducidos en la p. 577. Como $b_0 = 1$ y $b_n = 0$ para n impar, se tiene que

$$\begin{aligned} \frac{e^z + 1}{e^z - 1} \cdot \frac{z}{2} &= \frac{z}{2} \cdot \frac{e^{z/2} + e^{-z/2}}{e^{z/2} - e^{-z/2}} \\ &= \frac{z}{2} \cdot \frac{2 \cosh \frac{1}{2}z}{2 \operatorname{senh} \frac{1}{2}z} \qquad (52) \\ &= \tfrac{1}{2}z \coth \tfrac{1}{2}z = \sum_{n=0}^{\infty} b_{2n} z^{2n} = \sum \frac{B_{2n}^*}{(2n)!} z^{2n}. \end{aligned}$$

En esta forma se obtiene la serie de Taylor para la cotangente hiperbólica, que se dio en la p. 578. Los coeficientes de Taylor están relacionados de manera muy sencilla con los números de Bernoulli. Hemos, pues, demostrado que el desarrollo es válido para todo $|z| < 2\pi$.

De manera semejante se obtiene la serie de Taylor para la cotangente ordinaria (trigonométrica). Para $|z| < 2\pi$ y $0 < t < 1$, la función generadora

$$G(t, z) = \sum_{n=0}^{\infty} (-1)^n \psi_{2n}(t) z^{2n} \tag{53}$$

se deriva dos veces y así se encuentra que G satisface la ecuación diferencial $d^2G/dt^2 + z^2G = 0$, cuya solución general es $G = a \cos(zt) + b \operatorname{sen}(zt)$, donde a y b no dependen de t pero sí posiblemente del parámetro z. Para determinar a y b se usan dos condiciones. La primera, $\int_0^1 G(t, z)\, dt = 1$, se encuentra por integración término a término. La segunda,

$$\lim_{t \to 0} \frac{dG(t, z)}{dt} = \tfrac{1}{2}z^2$$

para todo z, se encuentra por derivación término a término, para lo cual se usa el hecho de que si $n > 1$ es

$$\psi_{2n}'(0) = \psi_{2n-1}(0) = b_{2n-1} = 0.$$

Estas condiciones implican que

$$a = \frac{z}{2} \cot \frac{z}{2}, \qquad b = \frac{z}{2},$$

de manera que para $|z| < 2\pi$ y $0 < t < 1$ se tiene

$$G(t, z) = \frac{z}{2} \frac{\cos(zt - z/2)}{\operatorname{sen}(z/2)}.$$

Los detalles se dejan al lector como ejercicio.

Haciendo $t \to 0$ en esta fórmula, se obtiene el desarrollo de Taylor para la cotangente (ver p. 578) con $|z| < 2\pi$:

$$(54) \qquad G(0, z) = \sum_{n=0}^{\infty} (-1)^n b_{2n} z^{2n} = \tfrac{1}{2} z \cot \tfrac{1}{2} z.$$

c. La fórmula de la suma de Euler-Maclaurin

En la sección 5.4*b* obtuvimos la fórmula de Taylor por medio de sucesivas integraciones por partes. A continuación presentamos una derivación análoga para una fórmula famosa debida a Euler, en la que los polinomios de Bernoulli, o, mejor dicho, sus extensiones periódicas ,toman el lugar que antes correspondió a los polinomios $(t - b)^n/n!$. (Reemplazamos entonces a y b en la p. 467 por 0 y 1; esto siempre puede hacerse transformando la variable t en la variable $s = (t - a)/(b - a)$, por lo que el cambio no es esencial.)

En vez de comenzar con la relación

$$(55) \qquad f(1) - f(0) = \int_0^1 f'(t)\, dt,$$

lo que correspondería a la deducción anterior de la fórmula de Taylor, empezamos ahora con la relación

$$(56) \qquad \int_0^1 f(t)\, dt = \int_0^1 f(t)\psi_0(t)\, dt,$$

que conducirá a una mayor simetría. Por ser

$$\psi_0(t) = \psi_1'(t), \qquad \psi_1(+0) = -\tfrac{1}{2},$$

y $\psi_1(1 - 0) = \frac{1}{2}$, la fórmula de integración por partes,

$$\int_0^1 u\, dv = uv \Big|_0^1 - \int_x^1 v\, du,$$

con $u = f(t)$, $v = \psi_1(t)$, $f(0) = f_0$, $f(1) = f_1$ conduce a (ver también capítulo 3, p. 298):

$$\int_0^1 f(t)\, dt = \tfrac{1}{2}(f_0 + f_1) - \int_0^1 f'(t)\psi_1(t)\, dt,$$

o sea,

$$(57)\qquad \tfrac{1}{2}(f_0 + f_1) = \int_0^1 f(t)\,dt + \int_0^1 f'(t)\psi_1(t)\,dt,$$

que es una expresión explícita para la desviación de la suma del lado izquierdo respecto de la integral $\int_0^1 f(t)\,dt$.

Puesto que una fórmula correspondiente es válida para cualquier intervalo entre dos enteros sucesivos, debido a la periodicidad de $\psi_1(t)$ se obtiene de inmediato la ecuación

$$(58)\qquad \tfrac{1}{2}f_0 + f_1 + f_2 + \cdots f_{n-1} + \tfrac{1}{2}f_n = \int_0^n f(x)\,dx + \int_0^n f'(x)\psi_1(x)\,dx,$$

o bien, para cualquier intervalo $a \leq x \leq b$ con a y b enteros,

$$(58a)\qquad f_a + f_{a+1} + \cdots + f_{b-1}$$
$$= \int_a^b f(x)\,dx + \int_a^b f'(x)\psi_1(x)\,dx, -\tfrac{1}{2}(f_b - f_a).$$

En esta forma se obtiene una expresión exacta para la diferencia entre la suma del lado izquierdo (el área de los rectángulos inscritos, en el caso de una función creciente) y el primer término del lado derecho (el área bajo la curva). La ecuación (58a) constituye la expresión más sencilla de la *fórmula de la suma de Euler-Maclaurin.*

Es natural tratar de mejorar este resultado repitiendo la integración por partes. Integrando la expresión $\int_a^b f'(x)\psi_1(x)\,dx$ y haciendo $u = f'(x)$, $dv = \psi_1(x)\,dx$ se obtiene

$$\int_a^b f'(x)\psi_1(x)\,dx = f'(x)\psi_2(x)\Big|_a^b - \int_a^b f''(x)\psi_2(x)\,dx.$$

Ahora bien, puesto que

$$\psi_2(b) = \psi_2(a) = \psi_2(0) = b_2,$$

el primer término toma la forma

$$b_2[f'(b) - f'(a)];$$

y el segundo término puede integrarse por partes nuevamente, dando

$$-b_3[f''(b) - f''(a)] + \int_a^b f'''(x)\psi_3(x)\,dx.$$

Como $b_3 = 0$, el primer término de esta última expresión se anula, y si se repite la integración por partes resulta

$$b_4[f'''(b) - f'''(a)] - \int_a^b f''''(x)\psi_4(x)\,dx.$$

Prosiguiendo de esta manera hasta llegar a ψ_{2k} se obtiene la *forma general* de la *fórmula de la suma de Euler,*

$$(59)\quad f_a + f_{a+1} + \cdots + f_{b-1} = \int_a^b f(x)\,dx - \tfrac{1}{2}[f(b) - f(a)] + \sum_{n=1}^{k} b_{2n}[f^{(2n-1)}(b) - f^{(2n-1)}(a)] + R_k,$$

en la que el residuo R_k puede escribirse en una de las formas siguientes:

$$(60)\qquad R_k = -\int_a^b f^{(2k)}(x)\psi_{2k}(x)\,dx$$

o bien

$$(60a)\qquad R_k = \int_a^b f^{(2k+1)}(x)\psi_{2k+1}(x)\,dx.$$

d. Aplicaciones. Desarrollos asintóticos

Desarrollos convergentes. La fórmula de la suma de Euler puede aplicarse bajo condiciones diferentes. En primer lugar, si $R_k \to 0$ cuando $k \to \infty$, la serie infinita

$$\sum_{n=1}^{\infty} b_{2n}[f^{(2n-1)}(b) - f^{(2n-1)}(a)]$$

converge, y así la fórmula permite expresar en forma exacta la suma de la serie correspondiente. También proporciona desarrollos en series para determinadas funciones.

Desarrollos no convergentes. Como segunda posibilidad, y acaso de mayor importancia, considérese el caso en que el residuo R_k puede no tender a cero cuando $k \to \infty$. Entonces la serie anterior no necesariamente converge. Sin embargo, puede ocurrir que al principio los valores absolutos $|R_k|$ decrezcan al crecer k y que $|R_k|$ sea muy pequeño para valores de k escogidos adecuadamente, pero que posteriormente (para valores de k mayores) $|R_k|$ empiece a crecer marcadamente. En este caso la fórmula de suma puede ser una herramienta importante para fines de cálculo numérico. Aunque no es posible obtener una precisión arbitrariamente grande, como con las series convergentes, se puede calcular el valor del lado izquierdo con un error que en el peor de los casos es igual al mínimo valor de $|R_k|$, lo que a menudo constituye una aproximación excelente. Examinaremos a continuación ejemplos de ambos fenómenos.

Ejemplo. Funciones exponenciales. Consideremos primero la función $f(x) = e^{xz}$ para algún z fijo. Con $a = 0$ y $b = 1$ obtenemos, para cualquier número k, la relación

$$f_0 = \int_0^1 f(x)\,dx - \tfrac{1}{2}[f(1) - f(0)]$$

$$+ \sum_{n=1}^{k} b_{2n}[f^{(2n-1)}(b) - f^{(2n-1)}(a)] + R_k.$$

En consecuencia,

$$1 = \frac{e^z - 1}{z} - \tfrac{1}{2}(e^z - 1) + \sum_{n=1}^{k} b_{2n}z^{2n-1}(e^z - 1) + R_k$$

$$= \frac{e^z - 1}{z}\cdot\left[1 - \frac{z}{2} + \sum_{n=1}^{k} b_{2n}z^{2n}\right] + R_k,$$

donde

$$R_k = -\int_0^1 z^{2k}e^{zx}\psi_{2k}(x)\,dx.$$

Puesto que $|\psi_{2k}(x)| \leqslant 4/(2\pi)^{2k}$ (p. 639), se encuentra que

$$|R_k| \leq |z|^{2k}\cdot e^{|z|}\cdot\frac{4}{(2\pi)^{2k}} = 4e^{|z|}\left(\frac{|z|}{2\pi}\right)^{2k},$$

o sea, $R_k \to 0$, al menos para $|z| < 2\pi$. Por tanto, para estos valores de z puede hacerse que k crezca más allá de toda cota en la fórmula de la suma obteniéndose

(61) $$\frac{z}{e^z - 1} = 1 - \tfrac{1}{2}z + \sum_{n=1}^{\infty} b_{2n}z^{2n}$$

para la función $z/(e^z - 1)$. Esta fórmula ya ha sido obtenida por otros métodos (p. 638). Obsérvese que el intervalo de convergencia es nuevamente $|z| < 2\pi$.

e. Sumas de potencias. Fórmula de recurrencia para los números de Bernoulli

Un ejemplo aún más sencillo de una fórmula de suma de Euler convergente se presenta cuando la serie del lado derecho contiene sólo un número finito de términos, especialmente si $f(x)$ es un polinomio de grado $r \geq 1$, de manera que $f^{(r+1)}(x)$ se anule idénticamente. Escogemos $f(x) = x^r$, $a = 0$, $b = n$ y $k > \tfrac{1}{2}r$. Por simplicidad, introducimos de nuevo la sucesión $B_n{}^*$ de números de Bernoulli, definida antes (p. 638), en la forma $B_n{}^* = n!\,b_n$ para todo n. Observando que

$$B_0{}^* = 1,\ B_1{}^* = -\tfrac{1}{2},\ B_3{}^* = B_5{}^* = B_7{}^* = \cdots = B^*_{2n+1} = 0,$$

vemos que la fórmula de Euler (59) se convierte en

$$\begin{aligned}
1 + 2^r + 3^r + \cdots + (n-1)^r &= \int_0^n x^r\,dx + \sum_{\nu=1}^{r} \frac{B_\nu^*}{\nu!}\cdot(f^{(\nu-1)}(n) - f^{(\nu-1)}(0)) \\
&= \frac{n^{r+1}}{r+1} + \sum_{\nu=1}^{r} \frac{B_\nu^*}{\nu!} r(r-1)\cdots(r-\nu+2)n^{r-\nu+1} \\
&= \frac{1}{r+1}\left\{n^{r+1} + \sum_{\nu=1}^{r}\binom{r+1}{\nu} B_\nu^* n^{(r+1)-\nu}\right\} \\
&= \frac{1}{r+1}\left\{\sum_{\nu=0}^{r+1}\binom{r+1}{\nu} n^{(r+1)-\nu} B_\nu^* - B_{r+1}^*\right\}.
\end{aligned}$$

Esta fórmula puede escribirse simbólicamente en la forma

$$(62)\qquad 1 + 2^r + 3^r + \cdots + (n-1)^r = \frac{1}{r+1}\{(n+B^*)^{r+1} - B^{*r+1}\},$$

en la que el término que aparece dentro del paréntesis deberá desarrollarse formalmente haciendo uso del teorema del binomio, y cada una de las "potencias" B^{*k} se reemplazará por el correspondiente número de Bernoulli, B_k^*. Por ejemplo,

$$\begin{aligned}
1 + 2^2 + 3^2 + \cdots + (n-1)^2 &= \tfrac{1}{3}(n^3 + 3n^2B_1 + 3nB_2) \\
&= \tfrac{1}{6}(2n^3 - 3n^2 + n) \\
1 + 2^3 + 3^3 + \cdots + (n-1)^3 &= \tfrac{1}{4}n^2(n-1)^2
\end{aligned}$$

(ver p. 81).

Haciendo $n = 1$ la fórmula (62) se reduce a

$$\frac{1}{r+1}\{(1+B^*)^{r+1} - B^{*r+1}\} = 0,$$

o sea,

$$(62a)\qquad (1+B^*)^{r+1} = B^{*r+1} \quad \text{para todo } r = 1;$$

que es precisamente la fórmula de recurrencia para B_k^* dada en la p. 579.

f. Constante de Euler y serie de Stirling

Un ejemplo de la aplicación de la fórmula de Euler-Maclaurin en el segundo caso, o sea, cuando se tiene divergencia, está dado por la función $f(x) = 1/x$ con $a = 1$ y $b = n$. Por (58*a*) se ve que

$$\begin{aligned}
(63)\qquad 1 + \tfrac{1}{2} + \tfrac{1}{3} + \cdots + \frac{1}{n-1} &= \int_1^n \frac{dx}{x} + \frac{1}{2}\left(1 - \frac{1}{n}\right) - \int_1^n \frac{\psi_1(x)}{x^2}\,dx \\
&= \log n + \tfrac{1}{2} - \frac{1}{2n} - \int_1^n \frac{\psi_1(x)}{x^2}\,dx,
\end{aligned}$$

o sea,

$$1 + \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{4} + \cdots + \frac{1}{n} - \log n = \tfrac{1}{2} + \frac{1}{2n} - \int_1^n \frac{\psi_1(x)}{x^2}\,dx.$$

Cuando $n \to \infty$ la integral de la derecha converge, pues $|\psi_1(x)| \leqslant \tfrac{1}{2}$ para todo x. Así, el valor absoluto del integrando es siempre menor que el de la integral convergente $\int_1^\infty dx/x^2$. Resulta entonces la relación

$$(64) \qquad \lim_{n\to\infty}\left[\sum_{k=1}^{n} \frac{1}{k} - \log n\right] = \tfrac{1}{2} - \int_1^\infty \frac{\psi_1(x)}{x^2}\,dx = C$$

donde C es una constante definida, la constante de Euler, que se introdujo ya en la p. 542.

Tenemos entonces dos resultados: 1) La serie armónica crece con la misma rapidez que la función logaritmo (ambas divergen a infinito); y 2) existe una expresión explícita para la diferencia entre las dos,

$$\sum_{k=1}^{n} \frac{1}{k} - \log n - C = R_n = \frac{1}{2n} + \int_n^\infty \frac{\psi_1(x)}{x^2}\,dx.$$

Obsérvese que R_n se anula cuando $n \to \infty$ por lo menos en primer orden.

Haciendo $f(x) = \log x$, $a = 1$, $b = n$ en la fórmula (59) de p. 642, se obtiene una aplicación más importante. En este caso,

$$\log 1 + \log 2 + \cdots + \log (n-1) = n \log n - n + 1 - \tfrac{1}{2}\log n$$
$$- \sum_{m=1}^{k} b_{2m}(2m-2)!\left(1 - \frac{1}{n^{2m-1}}\right) + \int_1^n \frac{(2k)!}{x^{2k+1}}\psi_{2k+1}(x)\,dx.$$

Sumando $\log n$ a ambos lados de esta ecuación, resulta

$$(65) \qquad \log n! = (n + \tfrac{1}{2})\log n - n + c_k + \sum_{m=1}^{k} \frac{(2m-2)!}{n^{2m-1}} b_{2m} - r_k(n),$$

donde

$$c_k = 1 - \sum_{m=1}^{k} b_{2m}(2m-2)! + \int_1^\infty \frac{(2k)!}{x^{2k+1}}\psi_{2k+1}(x)\,dx$$

$$r_k(n) = \int_n^\infty \frac{(2k)!}{x^{2k+1}}\psi_{2k+1}(x)\,dx.$$

Las integrales impropias convergen para $k > 0$, puesto que las funciones $\psi_{2k+1}(x)$ son periódicas y, por consiguiente, acotadas para todo x (ver p. 326). El valor de la constante c_k puede encontrarse observando que cuando $n \to \infty$ es

$$c_k = \lim_{n\to\infty} \log\left(\frac{n!\,e^n}{n^{n+1/2}}\right)$$

De la fórmula de Stirling (14) de p. 521 (o directamente del producto de Wallis para π dado en la p. 300) se concluye que $c_k = \log\sqrt{2\pi}$. Si los números de Bernoulli b_{2m} se expresan en la forma b_{2m} como $(-1)^{m-1}B_m/(2m)!$ (ver fórmula (47), p. 636), se obtiene la llamada *serie de Stirling:*

$$\log\left(\frac{n!}{\sqrt{2\pi}\,n^{n+1/2}e^{-n}}\right) = \sum_{m=1}^{k} \frac{(-1)^{m-1}B_m}{2m(2m-1)n^{2m-1}} - r_k(n).$$

Esta fórmula es un refinamiento de la "fórmula de Stirling". Para cualquier entero positivo fijo, k, y n suficientemente grande, los términos de la suma se aproximan a cero según los órdenes de aproximación $1/n$, $1/n^3$, $1/n^5, \ldots, 1/n^{2k-1}$, respectivamente. El término residual, $r_k(n)$, se aproxima a cero con la misma rapidez que $1/n^{2k}$, pues $\psi_{2k+1}(x)$ es una función acotada. Así, para un valor de k fijo y un valor de n muy grande, cada término de la suma resulta ser muy grande comparado con los términos siguientes, y el residuo será más pequeño que todos los términos de la suma. Se obtiene entonces una fórmula de aproximación de la forma

$$(66)\qquad \log\left(\frac{n!}{\sqrt{2\pi}\,n^{n+1/2}e^{-n}}\right) = \frac{B_1}{1\cdot 2}\frac{1}{n} - \frac{B_2}{3\cdot 4}\frac{1}{n^3} + \frac{B_3}{5\cdot 6}\frac{1}{n^5} - + \cdots$$

$$= \frac{1}{12}\frac{1}{n} - \frac{1}{360}\frac{1}{n^3} + \frac{1}{1260}\frac{1}{n^5} - \frac{1}{1680}\frac{1}{n^7} + \frac{1}{1188}\frac{1}{n^9} - + \cdots$$

Este desarrollo no debe ser considerado, sin embargo, como una serie infinita convergente. Es sólo *asintóticamente* correcto, en el sentido de que si cortamos la serie después de un número dado de términos, digamos, k términos, el error r_k es pequeño comparado con todos los términos que se conservan, siempre que n sea suficientemente grande. En ningún caso es posible lograr que el error sea arbitrariamente pequeño tomando más y más términos, una vez que se ha fijado n. De hecho, la serie infinita (66) diverge, como puede deducirse fácilmente de la estimación dada en la p. 637 para los números de Bernoulli. Dado un valor de n grande, existe un número de términos de la serie que podrían usarse. Así, para un valor de n moderadamente grande se tiene la aproximación

$$n! \approx \sqrt{2\pi}n^{n+1/2}e^{n+1/12n},$$

y cuando n es muy grande, la fórmula

$$n! \approx \sqrt{2\pi}n^{n+1/2}e^{n+1/12n-1/360n^3}$$

proporciona una aproximación más precisa; etc.

PROBLEMAS

SECCION 8.1, página 588

1. El *período fundamental*, T, de una función periódica, f, se define como el extremo inferior ("inf") de los períodos positivos de f. Demostrar que:

a) Si $T \neq 0$, entonces T es un período.

b) Si $T \neq 0$, entonces cualquier otro período es un múltiplo entero de T.

c) Si $T = 0$ y f es continua en cualquier punto, entonces f es una función constante.

2. Demostrar que si f tiene períodos inconmensurables, T_1 y T_2, entonces el período fundamental, T, es cero. Dar un ejemplo de una función no constante cuyos períodos sean inconmensurables.

3. Supóngase que los períodos fundamentales de f y g son, respectivamente, a y b. Si a y b son conmensurables, por ejemplo $a/b = q/p$, donde p y q son enteros primos entre sí, demostrar entonces que, por ejemplo, $f + g$ puede tener un período fundamental de cualquier valor m/n, siendo $m = aq = bp$ y n cualquier número natural.

SECCION 8.5, página 613

1. Obtener la serie de Fourier para la función $f(x) = \pi x$ en el intervalo $0 \leq x \leq 1$, en forma de una serie de senos pura y de una serie de cosenos pura.

2. Mostrar cómo puede representarse una función definida en un intervalo acotado arbitrario por medio de una serie de Fourier.

3. Obtener el producto infinito para el coseno a partir de la relación

$$\cos \pi x = \frac{\operatorname{sen} 2\pi x}{2 \operatorname{sen} \pi x}.$$

4. Usando los productos infinitos para el seno y para el coseno, evaluar

a) $\dfrac{2}{1}\cdot\dfrac{2}{3}\cdot\dfrac{6}{5}\cdot\dfrac{6}{7}\cdot\dfrac{10}{9}\cdot\dfrac{10}{11}\cdot\dfrac{14}{13}\cdots$

b) $2\cdot\dfrac{2}{3}\cdot\dfrac{4}{3}\cdot\dfrac{8}{9}\cdot\dfrac{10}{9}\cdot\dfrac{14}{15}\cdot\dfrac{16}{15}.$

5. Expresar en fracciones parciales la cotangente hiperbólica.

6. Determínense las propiedades especiales de los coeficientes de Fourier de los desarrollos para funciones pares e impares tales que $f(x) = f(\pi - x)$.

SECCION 8.6, página 619

1. Investigar la convergencia del desarrollo de Fourier

$$\cos x + \frac{\cos 2x}{2} + \frac{\cos 3x}{3} + \cdots$$

de la función $-\log 2 \left| \operatorname{sen} \dfrac{x}{2} \right|$.

SECCION 8.7, página 623

1. Demostrar la relación de Parseval para una función lisa por trozos, f, que puede tener cierto número de discontinuidades.

APENDICE II.1, página 634

1. Demostrar que

$$\phi_n(t) = \frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}B_k^* t^{n-k}.$$

2. Demostrar que para $n > 1$ vale

$$\phi_n(t) = (-1)^n \phi_n(1-t).$$

3. Usando la expresión en fracciones parciales para la cotangente, desarrollar en una serie de potencias en x la función $\pi x \cot \pi x$. Comparando ésta con la serie dada en las pp. 640-641, demostrar que

$$\sum_{\nu=1}^{\infty}\frac{1}{\nu^{2m}} = (-1)^{m-1}\frac{(2\pi)^{2m}}{2\cdot(2m)!}B_{2m}^*.$$

4. Demostrar que

$$\sum_{\nu=1}^{\infty}\frac{1}{(2\nu-1)^{2m}} = \frac{(-1)^{m-1}(2^{2m}-1)\pi^{2m}}{2(2m)!}B_{2m}^*.$$

5. Demostrar que

$$\sum_{\nu=1}^{\infty}\frac{(-1)^{\nu}}{\nu^{2m}} = \frac{(-1)^{m}(2^{2m}-2)\pi^{2m}}{2\cdot(2m)!}B_{2m}^*.$$

6. Usando los productos infinitos para el seno y para el coseno, demostrar que

$$a)\ \log\left(\frac{\operatorname{sen} x}{x}\right) = -\sum_{\nu=1}^{\infty}\frac{(-1)^{\nu-1}2^{2\nu-1}B_{2\nu}^*}{(2\nu)!\,\nu}x^{2\nu};$$

$$b)\ \log\cos x = -\sum_{\nu=1}^{\infty}\frac{(-1)^{\nu-1}2^{2\nu-1}(2^{2\nu}-1)B_{2\nu}^*}{(2\nu)!\,\nu}x^{2\nu}.$$

7. Verificar que

$$a)\ \int_0^1 \frac{\log x}{1-x}\,dx = -\frac{\pi^2}{6};$$

$$b)\ \int_0^1 \frac{\log x}{1+x}\,dx = -\frac{\pi^2}{12}.$$

9

Ecuaciones diferenciales para los tipos más simples de vibraciones

En varias ocasiones anteriores hemos encontrado ecuaciones diferenciales, esto es, ecuaciones a partir de las cuales es posible determinar una función desconocida, y tales que involucran no sólo la función misma sino también sus derivadas.

El problema más simple de este tipo es el de encontrar la integral indefinida de una función dada $f(x)$, o sea, encontrar una función $y = F(x)$ que satisfaga la ecuación diferencial $y' - f(x) = 0$. En el capítulo 3, p. 244, se demostró que una ecuación de la forma $y' = \alpha y$ es satisfecha por una función exponencial $y = ce^{ax}$, y las funciones trigonométricas se caracterizaron por ecuaciones diferenciales (p. 331). Como se vio en el capítulo 4 (pp. 423-424), las ecuaciones diferenciales surgen en relación con los problemas de mecánica, y, en realidad, muchas ramas de las Matemáticas puras y la mayor parte de las Matemáticas aplicadas dependen de las ecuaciones diferenciales. En este capítulo, sin profundizar en la teoría general, consideraremos las ecuaciones diferenciales de los tipos más simples de vibraciones. Éstas no son solamente de interés teórico sino que también son sumamente importantes en las Matemáticas aplicadas.

Será conveniente recordar las siguientes ideas y definiciones generales. Por una *solución* de una ecuación diferencial entendemos una función que, cuando se substituye en la ecuación diferencial, satisface ésta "idénticamente"; esto es, "para todos los valores considerados de la variable independiente". A menudo se usa el término *"integral"* en vez del término *"solución"*. En primer lugar, debido a que el problema es más o menos una generalización del problema ordinario de integrar; y en segundo lugar, debido a que frecuentemente ocurre que la solución se encuentra en realidad por integración.

9.1 Problemas de vibración en Mecánica y en Física

a. Las vibraciones mecánicas más simples

En el capítulo 4 (p. 423) ya se han considera los tipos más simples de vibraciones mecánicas. En ese capítulo se consideró una partícula de masa m, que está en libertad de moverse sobre el eje x y que tiende a regresar siempre a su posición inicial, $x = 0$, bajo la acción de una fuerza restauradora. La magnitud de esta fuerza la consideramos proporcional al desplazamiento x y, en efecto, la hicimos igual a $-kx$, donde k es una constante positiva y el signo negativo expresa el hecho de que la fuerza siempre está dirigida hacia el origen. En esta ocasión supondremos que también está presente una fuerza de fricción y que tal fuerza es proporcional a la velocidad de la partícula y opuesta a ella. Dicha fuerza está entonces dada por una expresión de la forma $-r\dot{x}$, donde r es una constante de fricción positiva. Finalmente, supondremos también que la partícula está sujeta a la acción de una fuerza $f(t)$, que es una función del tiempo t. Entonces, por la ley fundamental de Newton, el producto de la masa m y la aceleración $\ddot{x}$ debe ser igual a la fuerza total, es decir, la fuerza elástica más la fuerza de fricción más la fuerza externa. Esto se expresa mediante la ecuación

$$m\ddot{x} + r\dot{x} + kx = f(t). \tag{1}$$

Esta ecuación gobierna el movimiento de la partícula. Si recordamos los ejemplos anteriores de ecuaciones diferenciales, tales como el problema de integración para $\dot{x} = dx/dt = f(t)$, cuya solución es $x = \int f(t)\,dt + c$, o bien la solución de la ecuación diferencial particular $m\ddot{x} + kx = 0$ de la p. 424, observaremos que estos problemas tienen un número infinito de soluciones diferentes. En el presente caso también encontraremos que existe un número infinito de soluciones, lo que se expresa de la siguiente manera: Es posible encontrar una *solución general* o *integral completa* $x(t)$ de la ecuación diferencial, que dependa no sólo de la variable independiente t sino también de dos parámetros arbitrarios, c_1 y c_2, llamados *constantes de integración*. Asignando valores especiales a estas constantes se obtiene una solución particular, y *toda* solución puede encontrarse dando valores especiales a dichas constantes.

Lo anterior es perfectamente comprensible (ver también p. 422). No puede esperarse que la ecuación diferencial por sí sola determine por completo el movimiento. Por el contrario, es razonable pedir que para un instante dado, digamos, en el tiempo $t = 0$, sea posible escoger la posición inicial, $x(0) = x_0$, y la velocidad inicial, $\dot{x}(0) = \dot{x}_0$ (brevemente, el *"estado" inicial*) de manera arbitraria. En otras palabras, en el tiempo $t = 0$ la partícula debe poder iniciar su movimiento con una posición y una velocidad completamente arbitrarias. Una vez que esto se logra, cabe esperar que el movimiento subsiguiente esté completamente determinado.

Las dos constantes arbitrarias c_1 y c_2 en la solución general permiten seleccionar la solución particular que satisfaga las condiciones iniciales. En la siguiente sección veremos que esto sólo puede realizarse de una manera.

Si no hay fuerzas externas presentes, es decir, si $f(t) = 0$, el movimiento se llama *movimiento libre*. La ecuación diferencial se dice entonces que es *homogénea*. Si $f(t)$ es diferente de 0 para todos los valores de t, decimos que el movimiento es *forzado* y que la ecuación diferencial es *no homogénea*. El término $f(t)$ en ocasiones se llama también *término de perturbación*.

b. Oscilaciones eléctricas

Un sistema mecánico del tipo sencillo que se ha descrito puede realizarse físicamente sólo de manera aproximada. Un ejemplo lo tenemos en el péndulo, siempre que sus oscilaciones sean pequeñas. Las oscilaciones de una aguja magnética, las del núcleo de un teléfono o del diafragma de un micrófono, y otras vibraciones mecánicas, pueden representarse con un buen grado de aproximación por sistemas como el descripto. Sin embargo, existe otro tipo de fenómenos que corresponden de manera más precisa a la ecuación diferencial (1). Este es el caso, por ejemplo, de un circuito eléctrico oscilatorio.

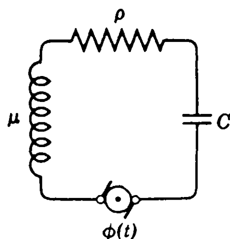


Figura 9.1 Circuito eléctrico oscilatorio.

Consideremos el circuito dibujado en la Fig. 9.1, con inductancia μ, resistencia ρ y capacitancia $C = 1/\kappa$. Supongamos que ese circuito está accionado por una fuerza electromotriz externa, $\phi(t)$, que es una función conocida del tiempo t, tal como el voltaje proporcionado por una dínamo o el voltaje producido por ondas eléctricas. Para describir el proceso que tiene lugar en el circuito denotemos el voltaje en el condensador por E y su carga por Q. Estas cantidades están entonces relacionadas por la ecuación $CE = E/\kappa = Q$. La corriente I, que, como el voltaje E, es una función del tiempo, se define como la rapidez de variación de la carga, por unidad de tiempo; es decir, como la rapidez con la cual disminuye la carga en el condensador: $I = -\dot{Q} = -dQ/dt = -\dot{E}/\kappa$. La ley de Ohm establece que el producto de la corriente por la resistencia es igual a la fuerza electromotriz (voltaje), o sea, es igual al voltaje del condensador, E, menos la fuerza contraelectromotriz debida a la autoinducción, más la fuerza electromotriz externa $\phi(t)$. En esta forma llegamos a la ecuación $I\rho = E - \mu\dot{I} + \phi(t)$, o $-(\rho/\kappa)\dot{E} = E + (\mu/\kappa)\ddot{E} + \phi(t)$, es decir, $\mu\ddot{E} + \rho\dot{E} + \kappa E = -\kappa\phi(t)$, que el voltaje en el circuito satisface. Vemos, por consiguiente, que resulta una ecuación diferencial precisamente del tipo (1). En vez de la masa, tenemos la inductancia; en vez de la fuerza de fricción, la resistencia; y en vez de la constante elástica, la inversa de la capacidad; en

tanto que la fuerza electromotriz externa (excepto por un factor constante) corresponde a la fuerza externa. Si la fuerza electromotriz es 0, la ecuación diferencial es homogénea.

Multiplicando ambos lados de la ecuación diferencial por $-1/\kappa$ y derivando con respecto al tiempo se obtiene para la corriente I la ecuación equivalente:

$$\mu\ddot{I} + \rho\dot{I} + \kappa I = \phi(t),$$

que difiere de la ecuación para el voltaje solamente por los términos del lado derecho, y que cuando se trata de oscilaciones libres ($\phi = 0$) tiene una forma idéntica.

9.2 Solución de la ecuación homogénea. Oscilaciones libres

a. La solución formal

Es fácil obtener una solución de la ecuación homogénea (1), $m\ddot{x} + r\dot{x} + kx = 0$, en la forma de una función exponencial, determinando una constante de manera tal que la expresión $e^{\lambda t} = x$ sea una solución, Si esta solución de prueba y sus derivadas $\dot{x} = \lambda e^{\lambda t}$, $\ddot{x} = \lambda^2 e^{\lambda t}$ se substituyen en la ecuación diferencial y se cancela luego el factor común $e^{\lambda t}$, se obtiene la ecuación cuadrática:

(2) $$m\lambda^2 + r\lambda + k = 0$$

para λ. Las raíces de esta ecuación son

$$\lambda_1 = -\frac{r}{2m} + \frac{1}{2m}\sqrt{r^2 - 4mk}, \qquad \lambda_2 = -\frac{r}{2m} - \frac{1}{2m}\sqrt{r^2 - 4mk}.$$

Cada una de las expresiones $x = e^{\lambda_1 t}$ y $x = e^{\lambda_2 t}$ es, al menos formalmente, una solución particular de la ecuación diferencial, como puede verse realizando las operaciones en sentido inverso. Pueden presentarse tres casos diferentes:

1. $r^2 - 4mk > 0$. Las dos raíces λ_1 y λ_2 son en este caso reales, negativas y diferentes, y la ecuación diferencial tendrá dos soluciones:

$$u_1 = e^{\lambda_1 t} \quad \text{y} \quad u_2 = e^{\lambda_2 t}.$$

Con ayuda de estas dos soluciones es posible construir de inmediato una solución en la que aparezcan dos constantes arbitrarias. En efecto, después de derivar vemos que

(3) $$x = c_1 u_1 + c_2 u_2$$

es también una solución de la ecuación diferencial. En la sección 9.3 demostraremos que esta expresión es, de hecho, la solución más general de la

ecuación; es decir, a partir de ella podemos obtener *cualquier* solución de la ecuación substituyendo para ello c_1 y c_2 por valores numéricos adecuados.

2. $r^2 - 4mk = 0$. La ecuación cuadrática tiene una raíz doble. Así se tiene inicialmente, excepto por un factor constante, sólo una solución: $x = w_1 = e^{-rt/2m}$. Pero en este caso es fácil verificar que la función

$$x = w_2 = te^{-rt/2m}$$

es también una solución de la ecuación diferencial.[1] En efecto, se encuentra que

$$\dot{x} = \left(1 - \frac{r}{2m}t\right)e^{-rt/2m}, \qquad \ddot{x} = \left(\frac{r^2}{4m^2}t - \frac{r}{m}\right)e^{-rt/2m},$$

y por substitución vemos que la ecuación diferencial

$$m\ddot{x} + r\dot{x} + \frac{r^2}{4m}x = m\ddot{x} + r\dot{x} + kx = 0$$

se satisface. Por tanto, la expresión

(4) $$x = c_1e^{-rt/2m} + c_2te^{-rt/2m}$$

da también una solución de la ecuación diferencial, con dos constantes de integración arbitrarias c_1 y c_2.

3. $r^2 - 4mk < 0$. Haciendo $r^2 - 4mk = -4m^2\nu^2$ obtenemos dos soluciones complejas de la ecuación diferencial, dadas por las expresiones $x = u_1 = e^{-rt/2m + i\nu t}$ y $x = u_2 = e^{-rt/2m - i\nu t}$. La fórmula de Euler

$$e^{\pm i\nu t} = \cos \nu t \pm i \operatorname{sen} \nu t$$

nos da, por una parte, las siguientes expresiones para las partes real e imaginaria de la solución compleja u_1:

$$v_1 = e^{-rt/2m}\cos \nu t, \qquad v_2 = e^{-rt/2m}\operatorname{sen} \nu t;$$

y, por otra parte, la representación

$$v_1 = \frac{u_1 + u_2}{2}, \qquad v_2 = \frac{u_1 - u_2}{2i}.$$

Por la segunda forma de la representación, vemos que v_1 y v_2 son soluciones (reales) de la ecuación diferencial. Verificar esto directamente por derivación y substitución es un ejercicio muy sencillo.

A partir de las dos soluciones particulares podemos construir una solución general,

(5) $$x = c_1v_1 + c_2v_2 = (c_1 \cos \nu t + c_2 \operatorname{sen} \nu t)e^{-rt/2m},$$

---

[1] Obtenemos esta solución, en forma natural, por medio del siguiente proceso de límite: si $\lambda_1 \neq \lambda_2$, entonces la expresión $(e^{\lambda_1 t} - e^{\lambda_2 t})/(\lambda_1 - \lambda_2)$ también representa una solución. Haciendo ahora λ_1 tender a λ_2 y escribiendo λ en vez de λ_1, λ_2, nuestra expresión se convierte en $d(e^{\lambda t})/d\lambda = te^{\lambda t}$.

con dos constantes arbitrarias, c_1 y c_2. Lo anterior puede también escribirse en la forma

(6) $$x = ae^{-rt/2m} \cos \nu(t - \delta),$$

donde se ha escrito $c_1 = a \cos \nu\delta$, $c_2 = a \operatorname{sen} \nu\delta$, y donde a, δ son dos nuevas constantes.

Como se recordará, esta solución ya ha sido obtenida antes para el caso especial $r = 0$ (sección 5.4).

b. Interpretación de la solución

En los casos $r > 2\sqrt{mk}$ y $r = 2\sqrt{mk}$ la solución está dada por la curva exponencial o por la gráfica de la función $te^{-rt/2m}$, que para valores grandes de t se asemeja a la curva exponencial, y también por la superposición de dichas curvas. En estos casos el proceso es aperiódico, es decir, cuando el tiempo aumenta la "distancia" x se aproxima asintóticamente a cero sin haber oscilaciones en la proximidad del valor $x = 0$. El movimiento no es, por tanto, oscilatorio. El efecto de la fricción o *amortiguamiento* es tan grande que impide que la fuerza elástica produzca movimientos oscilatorios.

Algo completamente diferente ocurre para $r < \sqrt{2mk}$, en cuyo caso el amortiguamiento es tan pequeño que aparecen raíces complejas, λ_1, λ_2. La expresión $x = a \cos \nu(t - \delta)e^{-rt/2m}$ da entonces *oscilaciones armónicas amortiguadas*. Son estas oscilaciones que siguen la ley sinusoidal y que tienen una frecuencia angular $\nu = \sqrt{k/m - r^2/4m^2}$, pero cuya amplitud, en lugar de ser constante, está dada por la expresión $ae^{-rt/2m}$. Esto es, la amplitud disminuye exponencialmente; y mientras mayor sea la expresión $r/2m$, más rápido será el decrecimiento. En la literatura de la Física este factor de amortiguamiento se llama a menudo la *constante de atenuación* de la oscilación amortiguada. El término indica que el logaritmo de la amplitud decrece con la razón de decrecimiento $r/2m$. En la Fig. 9.2 se ilustra una oscilación amortiguada de esta clase. Como hicimos antes, llamaremos a la cantidad $T = 2\pi/\nu$ *período* de la oscilación y a la cantidad $\nu\delta$ el *corrimiento de fase*. Para el caso especial $r = 0$ se obtienen oscilaciones armónicas simples con frecuencia $\nu_0 = \sqrt{k/m}$, la llamada *frecuencia natural* del sistema oscilatorio no amortiguado.

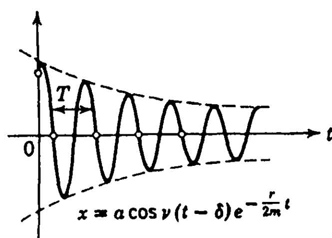


Figura 9.2 Oscilación armónica amortiguada.

c. Satisfacción de condiciones iniciales dadas. Unicidad de la solución

Queda todavía por demostrar que la solución en la que aparecen las dos constantes c_1 y c_2 puede hacerse corresponder a cualquier estado inicial prefijado, y también que tal solución representa todas las soluciones posibles de la ecuación. Supongamos que se ha encontrado una solución que en el tiempo $t = 0$ satisface las condiciones iniciales $x(0) = x_0$ $\dot{x}(0) = \dot{x}_0$, donde los números x_0 y $\dot{x}_0$ pueden tener cualquier valor. Entonces en el caso 1 de la sección 9.2*a* (p. 652) debemos hacer

$$c_1 + c_2 = x_0,$$

$$c_1\lambda_1 + c_2\lambda_2 = \dot{x}_0.$$

Tenemos así dos ecuaciones lineales para las constantes c_1 y c_2, y dichas ecuaciones admiten como soluciones únicas los valores

$$c_1 = \frac{\dot{x}_0 - \lambda_2 x_0}{\lambda_1 - \lambda_2}, \qquad c_2 = \frac{\dot{x}_0 - \lambda_1 x_0}{\lambda_2 - \lambda_1}.$$

En el caso 2 el mismo procedimiento da las dos ecuaciones lineales

$$c_1 = x_0,$$

$$\lambda c_1 + c_2 = \dot{x}_0 \left(\lambda = -\frac{r}{2m} \right),$$

a partir de las cuales es posible obtener nuevamente c_1 y c_2. Por último, en el caso 3 las ecuaciones que determinan las constantes toman la forma

$$a \cos \nu\delta = x_0,$$

$$a\left(\nu \operatorname{sen} \nu\delta - \frac{r}{2m} \cos \nu\delta \right) = \dot{x}_0,$$

cuyas soluciones son:

$$\delta = \frac{1}{\nu} \operatorname{arc\,cos} \frac{x_0}{a}, \qquad a = \frac{1}{\nu} \sqrt{\left[\nu^2 x_0^2 + \left(\dot{x}_0 + \frac{r}{2m} x_0 \right)^2 \right]}.$$

En esa forma hemos demostrado que las soluciones generales pueden satisfacer condiciones iniciales arbitrarias. Falta aún demostrar que no existe otra solución. Para ello basta probar que dado cierto estado inicial no existen dos soluciones diferentes.

Supongamos que existen dos soluciones, $u(t)$ y $v(t)$, tales que $u(0) = x_0$, $u(0) = \dot{x}_0$ y $v(0) = x_0$, $\dot{v}(0) = \dot{x}_0$. Entonces la diferencia es también una solución, por lo que se tendrá $w = u - v$. Esta solución corresponde, por consiguiente, a un estado inicial de reposo, es decir, a un estado tal que en

el tiempo $t = 0$ la partícula se encuentra en su posición de reposo y tiene velocidad cero. Pero en este caso la partícula nunca se pondría en movimiento. Para demostrarlo multipliquemos ambos lados de la ecuación diferencial $m\ddot{w} + r\dot{w} + kw = 0$ por $2\dot{w}$ y recordemos que $2\dot{w}\ddot{w} = (d/dt)\dot{w}^2$ y $2w\dot{w} = (d/dt)w^2$. Obtenemos así

$$\frac{d}{dt}(m\dot{w}^2) + \frac{d}{dt}(kw^2) + 2r\dot{w}^2 = 0.$$

Integrando entre los límites $t = 0$ y $t = \tau$ y usando la condición inicial $w(0) = 0$, $\dot{w}(0) = 0$, se encuentra que

$$m\dot{w}^2(\tau) + kw^2(\tau) + 2\dot{r}\int_0^\tau \left(\frac{dw}{dt}\right)^2 dt = 0.$$

Esta ecuación, sin embargo, encierra una contradicción cuando la función w es diferente de 0 para cualquier instante $\tau > 0$, pues el lado izquierdo es entonces positivo (ya que m, k y r son positivos) en tanto que el lado derecho es cero. Por tanto, $w = u - v$ es siempre igual a 0; lo que prueba que la solución es única.

9.3 La ecuación no homogénea. Oscilaciones forzadas

a. Comentarios generales. Superposición

Antes de proceder a la solución del problema en el caso de una fuerza externa $f(t)$, es decir, a la solución de la ecuación no homogénea, hacemos el siguiente comentario.

Si w y v son dos soluciones de la ecuación no homogénea, la diferencia $u = w - v$ satisface la ecuación homogénea. Esto puede verse de inmediato por substitución. Recíprocamente, si u es una solución de la ecuación homogénea y v lo es de la ecuación no homogénea, entonces $w = u + v$ es también una solución de esta última ecuación. Por tanto, a partir de una solución [1] de la ecuación no homogénea obtenemos *todas* las soluciones de dicha ecuación sumando la integral completa de la ecuación homogénea. Sólo necesitamos, por consiguiente, encontrar *una* solución de la ecuación no homogénea. Físicamente esto significa que si tenemos una oscilación forzada debido a una fuerza externa, al superponerle una oscilación libre arbitraria, representada por una solución de la ecuación homogénea, obtenemos un fenómeno que satisface la misma ecuación no homogénea que la oscilación forzada original. Si se trata de una fuerza de fricción, el movimiento libre en el caso oscilatorio debe tender a desaparecer a medida que el tiempo transcurre, debido al factor de amortiguamiento $e^{-rt/2m}$. Por

[1] Llamada a menudo *integral o solución particular.*

consiguiente, cuando se tiene una vibración forzada y existe una fuerza de fricción, es indiferente cuál sea la vibración libre superpuesta: el movimiento tenderá siempre hacia el mismo estado final al transcurrir el tiempo.

En segundo lugar, observamos que el efecto de una fuerza $f(t)$ puede partirse de la misma manera que la fuerza misma. Con esto queremos decir lo siguiente: si $f_1(t)$, $f_2(t)$ y $f(t)$ son tres funciones tales que

$$f_1(t) + f_2(t) = f(t),$$

y si $x_1 = x_1(t)$ es una solución de la ecuación diferencial $m\ddot{x} + r\dot{x} + kx = f_1(t)$, y $x_2 = x_2(t)$ es una solución de $m\ddot{x} + r\dot{x} + kx = f_2(t)$, entonces $x(t) = x_1(t) + x_2(t)$ es una solución de la ecuación

$$m\ddot{x} + r\dot{x} + kx = f(t). \tag{7}$$

Un resultado semejante se tiene, por supuesto, cuando $f(t)$ está compuesta por cualquier número de términos. Este resultado simple pero importante se llama *principio de superposición.* Su demostración es casi inmediata a partir de la ecuación misma. Subdividiendo la función $f(t)$ en dos o más términos es posible, de la misma manera, dividir la ecuación diferencial en varias ecuaciones, cuyo manejo, en determinadas circunstancias, puede resultar más sencillo.

El caso más importante es el de una fuerza externa, $f(t)$, periódica. Tal fuerza externa puede descomponerse en componentes periódicas puras, mediante un desarrollo en serie de Fourier, y en esa forma[1] puede aproximarse con la precisión deseada por medio de la suma de un número finito de tales funciones. Es suficiente, por tanto, encontrar la solución de la ecuación diferencial bajo la hipótesis de que su lado derecho tiene la forma

$$a \cos \omega t \quad \text{o,} \quad b \operatorname{sen} \omega t,$$

donde a, b y ω son constantes arbitrarias.

En vez de trabajar con estas funciones trigonométricas, podemos obtener la solución de manera más sencilla y elegante si usamos la notación compleja. Haciendo $f(t) = ce^{i\omega t}$, el principio de superposición muestra que sólo necesitamos considerar la ecuación diferencial

$$m\ddot{x} + r\dot{x} + kx = ce^{i\omega t}, \tag{8}$$

en la que c representa una constante arbitraria real o compleja. Tal ecuación diferencial representa en realidad dos ecuaciones diferenciales reales. Pues si partimos el lado derecho en dos términos, tomando, por ejemplo, $c = 1$, y escribimos $e^{i\omega t} = \cos \omega t + i \operatorname{sen} \omega t$, entonces x_1 y x_2, las soluciones de las dos ecuaciones diferenciales *reales* $m\ddot{x} + r\dot{x} + kx = \cos \omega t$ y $m\ddot{x} +$

[1] Supoı ndo que sea continua y lisa por trozos (p. 621), el caso más importante en Física.

$r\dot{x} + kx = \operatorname{sen} \omega t$, se combinan para dar la solución $x = x_1 + x_2$ de la ecuación diferencial *compleja*. Recíprocamente, resolviendo primero la ecuación diferencial en forma compleja, la parte real de su solución nos dará la función x_1, y la parte imaginaria la función x_2.

b. Solución de la ecuación no homogénea

Resolveremos la ecuación (8) mediante un artificio que surge en forma natural de la intuición. Supongamos que c es real y (para determinado instante) que $r \neq 0$. Adivinamos entonces que existirá un movimiento cuyo ritmo es el mismo que el de la fuerza periódica externa, y de acuerdo con esta hipótesis intentamos encontrar una solución de la ecuación diferencial, en la forma

$$(9) \qquad x = \sigma e^{i\omega t},$$

en la que sólo falta determinar el factor σ, que es independiente del tiempo. Substituyendo esta expresión y sus derivadas, $\dot{x} = i\omega\sigma e^{i\omega t}$, $\ddot{x} = -\omega^2\sigma e^{i\omega t}$, en la ecuación diferencial y eliminando el factor común $e^{i\omega t}$ obtenemos la ecuación

$$-m\omega^2\sigma + ir\omega\sigma + k\sigma = c,$$

o sea,

$$(10) \qquad \sigma = \frac{c}{-m\omega^2 + ir\omega + k}.$$

Recíprocamente, observamos que para este valor de σ la expresión es realmente una solución de la ecuación diferencial. Sin embargo, para expresar con claridad lo que este resultado significa es necesario realizar algunas transformaciones.

Comenzaremos escribiendo el factor complejo σ en la forma

$$(11) \qquad \sigma = c\,\frac{k - m\omega^2 - ir\omega}{(k - m\omega^2)^2 + r^2\omega^2} = c\alpha e^{-i\omega\delta},$$

donde el "factor de distorsión" positivo α y el "corrimiento de fase" $\omega\delta$ son expresados en términos de las cantidades dadas m, r, k mediante las ecuaciones

$$\alpha^2 = \frac{1}{(k - m\omega^2)^2 + r^2\omega^2}, \qquad \operatorname{sen} \omega\delta = r\omega\alpha, \qquad \cos \omega\delta = (k - m\omega^2)\alpha.$$

Con esta notación nuestra solución toma la forma

$$x = c\alpha e^{i\omega(t-\delta)},$$

cuyo significado es el siguiente: a la fuerza $c \cos \omega t$ le corresponde el "efecto" $c\alpha \cos \omega(t - \delta)$, y a la fuerza $c \operatorname{sen} \omega t$ le corresponde el efecto $c\alpha \operatorname{sen} \omega(t - \delta)$.

Vemos por lo anterior que el efecto es una función del mismo tipo que la fuerza, esto es, una *oscilación* no amortiguada (sostenida). Esta oscilación difiere de la oscilación que representa a la fuerza en que su amplitud ha aumentado en la proporción $\alpha : 1$ y la fase ha variado según el ángulo $\omega\delta$. Por supuesto, el mismo resultado puede obtenerse fácilmente sin hacer uso de la notación compleja, aunque a costa de cálculos algo más complicados.

De acuerdo con el comentario que se hizo al principio de esta sección, una vez encontrada esta solución tenemos completamente resuelto el problema; pues superponiendo cualquier oscilación libre podemos obtener la oscilación forzada más general.

Reuniendo los resultados anteriores podemos establecer lo siguiente:

La integral completa[1] *de la ecuación diferencial*

$$m\ddot{x} + r\dot{x} + kx = ce^{i\omega t}$$

(*en la que* $x \neq 0$) *es* $x = c\alpha e^{i\omega(t-\delta)} + u$, *donde* u *es la integral completa de la ecuación homogénea* $m\ddot{x} + r\dot{x} + kx = 0$, y las cantidades α y δ están definidas por las ecuaciones

$$\alpha^2 = \frac{1}{(k - m\omega^2)^2 + r^2\omega^2}; \qquad \operatorname{sen}\omega\delta = r\omega\alpha, \qquad \cos\omega\delta = (k - m\omega^2)\alpha.$$

Las constantes de esta solución general permiten hacer que la solución satisfaga un estado inicial arbitrario, es decir, dados cualesquiera valores de x_0 y $\dot{x}_0$, las constantes pueden escogerse de manera tal que $x(0) = x_0$ y $\dot{x}(0) = \dot{x}_0$.

c. La curva de resonancia

Para formarnos una idea de la solución que hemos obtenido y de lo que significa en las aplicaciones, estudiaremos el "factor de distorsión", α, como función de la "frecuencia de excitación", ω, es decir, la función

$$(13) \qquad \phi(\omega) = \frac{1}{\sqrt{(k - m\omega^2)^2 + r^2\omega^2}}.$$

Tal investigación en detalle es motivada por el hecho de que, dadas las constantes k, m y r o, como a veces se dice, dado un "sistema oscilatorio", es posible considerar éste como sujeto a fuerzas de excitación periódicas cuyas frecuencias angulares son muy diferentes entre sí; siendo importante, por tanto, considerar la solución de la ecuación diferencial para este amplio rango de fuerzas de excitación. Con el fin de describir en forma conveniente la función, introducimos la cantidad $\omega_0 = \sqrt{k/m}$. Este número, ω_0, es la

[1] O *general*.

frecuencia angular que tendrían las oscilaciones libres del sistema si la fricción fuese nula; o, más brevemente, la *frecuencia (angular) natural del sistema no amortiguado* (ver p. 655). En realidad, la frecuencia del sistema *libre,* en presencia de la función r, no es igual a ω_0 sino que tiene el valor

$$\nu = \sqrt{\frac{k}{m} - \frac{r^2}{4m^2}},$$

donde suponemos que $4km - r^2 > 0$. (Si éste no es el caso, el sistema libre no tiene frecuencia: es aperiódico.)

La función $\phi(\omega)$ tiende asintóticamente al valor cero a medida que la frecuencia de excitación tiende a infinito y, de hecho, tal función se anula con orden $1/\omega^2$. Además, $\phi(0) = 1/k$; en otras palabras, una fuerza excitadora de frecuencia cero y magnitud igual a 1, o sea, una fuerza constante de magnitud uno, da origen a un desplazamiento $1/k$ del sistema oscilatorio. En la región de valores positivos de ω la derivada $\phi'(\omega)$ no puede anularse excepto allí donde se anula la derivada de la expresión $(k - m\omega^2)^2 + r^2\omega^2$, esto es, para un valor $\omega = \omega_1 > 0$ que satisfaga la ecuación

$$-4m\omega(k - m\omega^2) + 2r^2\omega = 0.$$

Es claro que para que tal valor exista deberá tenerse que $2km - r^2 > 0$; y en este caso

$$\omega_1 = \sqrt{\frac{k}{m} - \frac{r^2}{2m^2}} = \sqrt{\omega_0^2 - \frac{r^2}{2m^2}}.$$

Puesto que la función $\phi(\omega)$ es positiva en todas partes, crece monótonamente para valores pequeños de ω y se anula en el infinito, el valor ω_1 debe dar un máximo. Llamamos a esta frecuencia ω_1 la "frecuencia de resonancia" del sistema.

Substituyendo la expresión para ω_1 encontramos que el valor del máximo es

$$\phi(\omega_1) = \frac{1}{r\sqrt{(k/m - r^2/4m^2)}}.$$

Cuando $r \rightarrow 0$ dicho valor máximo crece más allá de toda cota. Para $r = 0$, es decir, para un sistema oscilatorio no amortiguado, la función $\phi(\omega)$ tiene una discontinuidad infinita en el valor $\omega = \omega_1$. Es éste un caso límite, que consideraremos más adelante en forma especial.

La gráfica de la función $\phi(\omega)$ se llama *curva de resonancia* del sistema. Para $\omega = \omega_1$ (y, en consecuencia, para valores pequeños de r en la vecindad de la frecuencia "natural") la distorsión de amplitud, $\alpha = \phi(\omega)$, es particularmente grande. Este resultado es la expresión matemática del "fenómeno de resonancia", el cual, para valores fijos de m y k, es más evidente a medida que r decrece.

En la Fig. 9.3 se muestra una familia de curvas de resonancias correspondientes, todas ellas, a los valores $m = 1$ y $k = 1$ (y, por tanto, a $\omega_0 = 1$), pero con valores $D = \frac{1}{2}r$ diferentes. Observamos que para valores pequeños de D aparece una resonancia muy marcada cerca de $\omega = 1$; y en el caso límite, $D = 0$, $\phi(\omega)$ tendría para $\omega = 1$ una discontinuidad infinita en vez de un máximo. Al crecer D, el máximo se mueve hacia la izquierda, y para el valor $D = 1/\sqrt{2}$ tenemos $\omega_1 = 0$. En este último caso el punto en el que la tangente es horizontal se ha movido hacia el origen y el máximo ha desaparecido. Si $D > 1/\sqrt{2}$ el cero de $\phi'(\omega)$ no existe, y la curva de resonancia deja de tener máximo y la resonancia desaparece.

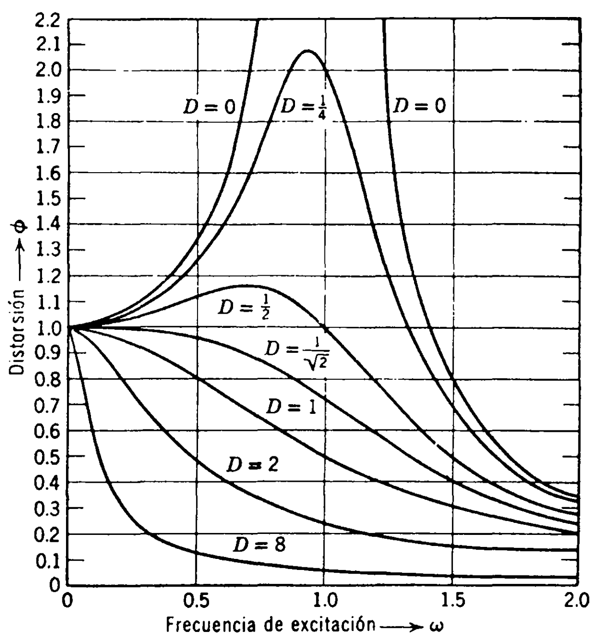


Figura 9.3 Curvas de resonancia.

En general, el fenómeno de resonancia desaparece tan pronto como se cumple la condición

$$2km - r^2 \leq 0.$$

En el caso de la igualdad, la curva de resonancia alcanza su máxima altura, $\phi(0) = 1/k$ para $\omega_1 = 0$; y su tangente es horizontal allí; y la curva, después de un camino inicialmente casi horizontal, decrece hacia el valor cero.

d. Una discusión adicional sobre las oscilaciones

La discusión anterior no es suficiente. Para entender en realidad el fenómeno del movimiento forzado es necesario hacer hincapié en lo siguiente. La integral particular $c\alpha e^{i\omega(t-\delta)}$ debe considerarse como un *estado límite* hacia el cual tiende cada vez más, a medida que el tiempo transcurre, la integral completa

$$x(t) = c\alpha e^{i\omega(t-\delta)} + c_1u_1 + c_2u_2.$$

Esto es así porque la oscilación libre $c_1u_1 + c_2u_2$, superpuesta a la integral particular, se desvanece al transcurrir el tiempo. Este desvanecimiento será lento si r es pequeño y será rápido si r es grande.

Supongamos, por ejemplo, que al iniciarse el movimiento, esto es, en el tiempo $t = 0$, el sistema está en reposo, de manera que $x(0) = 0$ y $\dot{x}(0) = 0$. Con estos datos podemos determinar las constantes c_1 y c_2, las cuales, como puede verse de inmediato, *no* se anulan simultáneamente. Aunque la frecuencia de excitación es aproximada o exactamente igual a ω_1, de manera que existe resonancia, la relativamente gran amplitud $\alpha = \phi(\omega_1)$ no aparece al principio. Por el contrario, quedará encubierta por la función $c_1u_1 + c_2u_2$ y aparecerá por primera vez cuando aquélla se haya desvanecido; es decir, aparece más lentamente cuando r es menor.

Para el sistema no amortiguado, esto es, para $r = 0$, la solución "falla" si la frecuencia de excitación es igual a la frecuencia natural $\omega_0 = \sqrt{k/m}$, pues entonces $\phi(\omega_0)$ es infinita. Por consiguiente, no es posible en este caso obtener una solución de la ecuación $m\ddot{x} + kx = e^{i\omega t}$ en la forma $\sigma e^{i\omega t}$. Es posible, sin embargo, obtener inmediatamente una solución particular de la forma $x = \sigma t e^{i\omega t}$. Substituyendo esta expresión en la ecuación diferencial y recordando que

$$\dot{x} = \sigma e^{i\omega t}(1 + i\omega t), \qquad \ddot{x} = \sigma e^{i\omega t}(2i\omega - t\omega^2),$$

se tiene

$$\sigma(2im\omega - m\omega^2 t + kt) = 1,$$

y, puesto que $m\omega^2 = k$,

$$\sigma = \frac{1}{2im\omega}.$$

Así, *cuando existe resonancia en un sistema no amortiguado la solución es de la forma*

$$x = \frac{t}{2im\omega} e^{i\omega t} = \frac{t}{2i\sqrt{km}} e^{i\omega t}.$$

Usando la notación real, cuando $f(t) = \cos \omega t$ se tiene que

$$x = \frac{1}{2} \frac{t}{\sqrt{km}} \operatorname{sen} \omega t$$

y cuando $f(t) = \text{sen}\,\omega t$ el resultado es

$$x = -\frac{1}{2}\frac{t}{\sqrt{km}}\cos\omega t.$$

Hemos, pues, encontrado una función que podríamos llamar "oscilación", pero cuya amplitud crece linealmente con el tiempo. La oscilación libre superpuesta no se desvanece, pues no es amortiguada; sin embargo, conserva su amplitud original y llega a no ser importante en comparación con la amplitud creciente de la oscilación forzada especial. El hecho de que en este caso la solución oscile hacia arriba y hacia abajo entre cotas positivas y negativas que crecen continuamente con el tiempo representa el significado real de la discontinuidad infinita que presenta la función de resonancia para un sistema no amortiguado.

e. Comentarios sobre la construcción de instrumentos de registro

En una gran variedad de aplicaciones en la Física y en la Ingeniería resulta de suma importancia la discusión precedente. En el caso de muchos instrumentos, tales como galvanómetros, sismógrafos, circuitos osciladores de radiorreceptores y en diafragmas de micrófonos, el problema consiste en registrar un desplazamiento x producido por una fuerza externa periódica. En tales casos la cantidad x satisface nuestra ecuación diferencial, por lo menos en primera aproximación.

Si T es el período de oscilación de la fuerza externa, es posible desarrollar ésta en una serie de Fourier de la forma

$$f(t) = \sum_{l=-\infty}^{\infty} \gamma_l e^{il(2\pi/T)t},$$

o, mejor aún, considerarla como representada con suficiente aproximación por la suma trigonométrica $\sum_{l=-N}^{N} \gamma_l e^{il(2\pi/T)t}$ compuesta solamente por un número finito de términos. Por el principio de superposición (p. 657), la solución $x(t)$ de la ecuación diferencial estará representada (no considerando la oscilación libre superpuesta) por una serie infinita [1] de la forma

$$x(t) = \sum_{l=-\infty}^{\infty} \sigma_l e^{il(2\pi/T)t},$$

o aproximadamente por una expresión finita de la forma

$$x(t) = \sum_{l=-N}^{N} \sigma_l e^{il(2\pi/T)t}.$$

En vista de nuestros resultados anteriores,

$$\sigma_l = \gamma_l \alpha_l e^{-i\delta_l(2\pi l/T)}$$

y

$$\alpha_l^2 = \frac{1}{\left(k - ml^2\dfrac{4\pi^2}{T^2}\right)^2 + r^2l^2\dfrac{4\pi^2}{T^2}}, \quad \tan\frac{2\pi l}{T}\delta_l = \frac{2\pi lr}{T\left(k - m\dfrac{4\pi^2l^2}{T^2}\right)}$$

[1] Las cuestiones de convergencia no se discutirán aquí.

Podemos entonces describir la acción de cualquier fuerza externa periódica de la siguiente manera: si descomponemos la fuerza de excitación en componentes periódicas puras, los términos individuales de la serie de Fourier, entonces cada componente experimenta una distorsión de amplitud y un corrimiento de fase propios, y los efectos separados pueden luego superponerse por adición. Si sólo nos interesa la distorsión de amplitud (en las aplicaciones el corrimiento de fase es sólo de importancia secundaria [2] y además puede tratarse de la misma manera que la distorsión de amplitud), un estudio de las curvas de resonancia nos da plena información acerca de la forma en que los movimientos del aparato de registro "siguen" a la fuerza de excitación externa. Para valores muy grandes de l, o de $\omega[=(2\pi/T)l]$, el efecto de la frecuencia de excitación sobre el desplazamiento x difícilmente será perceptible. Por otra parte, *todas* las frecuencias de excitación en la vecindad de ω_1, la frecuencia (angular) de resonancia, afectarán marcadamente a la cantidad x.

En la construcción de aparatos de medición y registro de cantidades físicas, las constantes m, r y k están a nuestra disposición, por lo menos dentro de límites amplios. Estos valores podrían escogerse de manera que la forma de la curva de resonancia se adapte lo mejor posible a las necesidades especiales de las mediciones que interesan. Intervienen aquí dos consideraciones de importancia. En primer lugar, es deseable que el aparato tenga la máxima sensibilidad posible; esto es, que para todas las frecuencias ω en cuestión el valor de α sea tan grande como las condiciones lo permitan. Para valores pequeños de ω, como ya hemos visto, α es aproximadamente proporcional a $1/k$, de manera que el número $1/k$ es una medida de la sensibilidad del instrumento para frecuencias de excitación pequeñas. La sensibilidad puede, por consiguiente, mejorarse aumentando $1/k$, o sea, haciendo más débil la fuerza restauradora.

El otro punto importante se refiere a la necesidad de una *ausencia de distorsión relativa*, o "indiferencia" (con respecto a las frecuencias). Supongamos que la representación $f(t) = \sum_{l=-N}^{N} \gamma_l e^{il(2\pi/T)t}$ es la adecuada para la fuerza de excitación. Decimos entonces que el aparato registra la fuerza de excitación con ausencia de distorsión relativa o con "indiferencia" si para todas las frecuencias angulares $\omega \leq N(2\pi/T)$ el factor de distorsión tiene aproximadamente el mismo valor. Esta condición es indispensable cuando se desea obtener conclusiones acerca del proceso de excitación directamente a partir del comportamiento del aparato, por ejemplo, si se desean reproducir sonidos musicales tanto graves con o agudos por medio de un radio o una grabadora de cinta, y con una relación de intensidades aproximadamente correcta. Una reproducción "sin distorsión" nunca puede lograrse en forma total, pues la curva de resonancia no tiene porción horizontal alguna. Puede, sin embargo, intentarse escoger las constantes m, k y r del aparato de manera tal que no exista una marcada resonancia, y también para que la curva tenga una tangente horizontal en su comienzo, con lo cual $\varphi(\omega) = \alpha$ será aproximadamente constante para valores pequeños de ω. Como ya sabemos, esto puede lograrse haciendo

$$2km - r^2 = 0.$$

Dadas las constantes m y k, es posible reunir los requisitos mencionados ajustando adecuadamente el valor de la fricción, r, lo cual se logra, en el caso de un circuito eléctrico, por ejemplo, insertando una resistencia de valor conveniente. La curva de resonancia nos muestra que entonces el sistema no amortiguado prácticamente no introduce distorsión para frecuencias que van desde 0 hasta la frecuencia natural, ω_0; y que por encima de esta frecuencia el amortiguamiento es considerable. Obtenemos, por consiguiente, una ausencia de distorsión relativa en un intervalo dado de frecuencias escogiendo, primero, m tan pequeña y k tan grande que la frecuencia natural ω_0 del sistema no amortiguado sea mayor que cualquiera de las frecuencias de excitación consideradas, y luego escogiendo el factor de amortiguamiento, r, de acuerdo con la ecuación $2km - r^2 = 0$.

---

[2] Ya que, por ejemplo, es imperceptible por el oído humano.

Lista de fechas biográficas

Abel, Niels Henrik (1802-1829)
Archimedes (287?-212 a.C.)
Barrow, Isaac (1630-1677)
Bernoulli, Jakob (1654-1705)
Bernoulli, Johannes (1667-1748)
Bessel, Friedrich Wilhelm (1784-1846)
Bolzano, Bernhard (1781-1848)
Brahe, Tycho (1546-1601)
Briggs, Henry (1556?-1630)
Cantor, Georg (1845-1918)
Cauchy, Augustin (1789-1857)
Coulomb, Charles Augustin de (1736-1806)
Darboux, Gaston (1842-1917)
Dedekind, Richard (1831-1916)
De Moivre, Abraham (1667-1754)
Descartes, (Cartesius) René (1596-1650)
Dirichlet, Gustav Lejeune (1805-1859)
Einstein, Albert (1879-1955)
Euclid (alrededor de 300 a.C.)
Euler, Leonhard (1707-1783)
Fejer, Lipot (1880-1959)
Fermat, Pierre de (1601-1665)
Fourier, Joseph (1768-1830)
Fresnel, Augustin (1788-1827)
Gauss, Carl Friedrich (1777-1855)
Gibbs, Josiah Willard (1839-1903)
Gregory, James (1638-1675)
Guldin, Paul (1577-1643)
Hermite, Charles (1822-1901)

Hölder, Otto (1860-1937)
Huygens, Christian (1629-1695)
Jensen, J. L. W. V. (1859-1925)
Kepler, Johannes (1571-1630)
Lagrange, Joseph Louis (1736-1813)
Lambert, Johann Heinrich (1728-1777)
Landau, Edmund (1877-1938)
Leibnitz, Gottfried Wilhelm von (1646-1716)
L'Hôpital, Guillaume, François Antoine de (1661-1704)
Lipschitz, Rudolf Otto (1832-1903)
Lorentz, Hendrik Antoon (1853-1928)
Maclaurin, Colin (1698-1746)
Michelson, Albert Abraham (1852-1931)
Morley, Edmund Williams (1838-1923)
Napier, John (1550-1617)
Newton, Isaac (1642-1727)
Ohm, Georg Simon (1787-1854)
Parseval, Marc Anton (A. ?-1836)
Ptolemy (Claudius Ptolemaeus) (Siglo II a.C.)
Raabe, Joseph Ludwig (1801-1859)
Riemann, Bernhard (1826-1866)
Rolle, Michael (1652-1719)
Schwarz, Hermann Amandus (1843-1921)
Seidel, Philipp Ludwig von (1821-1896)
Simpson, Thomas (1710-1761)
Stirling, James (1692-1770)
Taylor, Brook (1685-1731)
Vega, George (1754-1802)
Wallis, John (1616-1703)
Weierstrass, Karl (1815-1897)

Indice

A

B

C

F

R

S

V

W

Z

—o0o—

La edicion, composición, diseño e impresión de esta obra fueron realizados
bajo la supervisión de GRUPO NORIEGA EDITORES
Balderas 95, Col. Centro. México, D.F. C.P. 06040
Tel.: (5)521 48 49 y (5)512 30 09 Fax: (5)512 29 03 y (5)510 94 15
Impreso en Programas Educativos, S.A. de C.V. • 117754 000 06 99 517
Certificado por ISO-9002:1994/NMX-CC-004:1995 con el No. de registro RSC-048
e-mail: limusa@noriega.com.mx • www.noriega.com.mx